(54) Title: DESIGN AND DISCOVERY OF PROTEIN BASED TNF-α VARIANTS FOR THE TREATMENT OF TNF-α RELATED DISORDERS

(57) Abstract: The invention relates to novel proteins with TNF-α antagonist activity and nucleic acids encoding these proteins. The invention further relates to the use of the novel proteins in the treatment of TNF-α related disorders, such as rheumatoid arthritis.

DESIGN AND DISCOVERY OF PROTEIN BASED TNF-α VARIANTS FOR THE TREATMENT OF TNF-α RELATED DISORDERS

This application is a continuing application of U.S.S.N. 60/186,427, filed March 2, 2000.

FIELD OF THE INVENTION

The invention relates to novel proteins with TNF-α antagonist activity and nucleic acids encoding these proteins. The invention further relates to the use of the novel proteins in the treatment of TNF-α related disorders, such as rheumatoid arthritis.

BACKGROUND OF THE INVENTION

Tumor necrosis factor α (TNF-α) is a pleiotropic cytokine that is primarily produced by activated macrophages and lymphocytes; but is also expressed in endothelial cells and other cell types. TNF-α is a major mediator of inflammatory, immunological, and pathophysiological reactions. (Grell, M., et al., (1995) Cell, 83:793-802). Two distinct forms of TNF exist, a 26 kDa membrane expressed form and the soluble 17 kDa cytokine which is derived frm proteolytic cleavage of the 26 kDa form. The soluble TNF polypeptide is 157 amino acids long and is the primary biologically active molecule.

TNF-α exerts its biological effects through interaction with high-affinity cell surface receptors. Two distinct membrane TNF-α receptors have been cloned and characterized. These are a 55 kDa species, designated p55 TNF-R and a 75 kDa species designated p75 TNF-R (Corcoran. A.E., et al., (1994) Eur. J. Biochem., 223:831-840). The two TNF receptors exhibit 28% similarity at the amino acid level. This is confined to the extracellular domain and consists of four repeating cysteine-rich motifs, each of approximately 40 amino acids. Each motif contains four to six cysteines in conserved positions. Dayhoff analysis shows greatest intersubunit similarity among the first three repeats in each receptor. This characteristic structure is shared with a number of other receptors and cell surface molecules which comprise the TNF-R/nerve growth factor receptor superfamily (Corcoran. A.E., et al., (1994) Eur. J. Biochem., 223:831-840).

TNF signaling is initiated by receptor clustering, either by the trivalent ligand TNF or by cross-linking monoclonal antibodies (Vandevoorde, V., et al., (1997) J. Cell Biol., 137:1627-1638). Crystallographic

studies of TNF and the structurally related cytokine, lymphotoxin (LT) have shown that both cytokines exist as homotrimers, with subunits packed edge to edge in a threefold symmetry. Structurally, neither TNF or LT reflect the repeating pattern of the their receptors. Each monomer is cone shaped and contains two hydrophilic loops on opposite sides of the base of the cone. Recent crystallization of a p55 soluble TNF-R/LT complex has confirmed the hypothesis that loops from adjacent monomers join together to form a groove between monomers and that TNF-R binds in these grooves (Corcoran. A.E., et al., (1994) Eur. J. Biochem., 223:831-840).

The key role played by TNF-α in inflammation, cellular immune responses and the pathology of many diseases has led to the search for antagonists of TNF-α. Soluble TNF receptors which interfere with TNF-α signaling have been isolated and are marketed by Immunex as Enbrel® for the treatment of rheumatoid arthritis. Random mutagenesis has been used to identify active sites in TNF-α responsible for the loss of cytotoxic activity (Van Ostade, X., et al., (1991) EMBO J., 10:827-836). However, a need still exists to develop more potent TNF-α antagonists for use as therapeutic agents.

Accordingly, it is an object of the invention to provide proteins with TNF-α antagonist activity and nucleic acids encoding these proteins for the treatment of TNF-α related disorders.

SUMMARY OF THE INVENTION

In accordance with the objects outlined above, the present invention provides non-naturally occurring variant TNF-α proteins (e.g. proteins not found in nature) comprising amino acid sequences with at least one amino acid change compared to the wild-type TNF-α proteins. Preferred embodiments utilize variant TNF-α proteins that preferentially interact with the wild-type TNF-α to form mixed trimers incapable of activating receptor signaling. Preferably, variant TNF-α proteins with 1, 2, 3, 4, and 5 amino acid changes are used as compared to wild-type TNF-α protein. In a preferred embodiment these changes are selected from positions 21, 30, 31, 32, 33, 35, 65, 66, 67, 111, 112, 115, 140, 143, 144, 146 and 147.

In an additional aspect, the non-naturally occurring variant TNF-α proteins have substitutions selected from the group of substitutions consisting of D143E, D143N, D143S, A145R, A145K, A145E, E146K, E146R and A84V.

In a further aspect, the invention provides recombinant nucleic acids encoding the non-naturally occurring variant TNF-α proteins, expression vectors, and host cells.

2

In an additional aspect, the invention provides methods of producing a non-naturally occurring variant TNF-α protein comprising culturing the host cell of the invention under conditions suitable for expression of the nucleic acid.

In a further aspect, the invention provides pharmaceutical compositions comprising a variant TNF-α protein of the invention and a pharmaceutical carrier.

In a further aspect, the invention provides methods for treating an TNF-α related disorder comprising administering a variant TNF-α protein of the invention to a patient.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the design strategy for TNF-α mutants. Figure 1A depicts a complex of TNF receptor with wild-type TNF-α. Figure 1B depicts a mixed trimer of mutant TNF-α (TNF-X) and wild-type TNF-α. Dark circles are receptor molecules, light pentagons are wild-type TNF-α and the dark pentagon is a mutant TNF-α.

Figure 2 depicts the structure of the wild-type TNF-TNF-R trimer complex.

Figure 3 depicts the structure of the p55 TNF-R extra-cellular domain. The darker appear regions represent residues required for contact with TNF-α.

Figure 4 depicts the binding sites on TNF-α that are involved in binding the TNF-R.

Figure 5 depicts the TNF-α trimer interface.

Figure 6A depicts the nucleotide sequence of the histidine tagged wild-type TNF-α molecule used as a template molecule form which the mutants were generated. The additional 6 histidines, located between the start codon and the first amino acid are underlined.

Figure 6B depicts the amino acid sequence of wild-type TNF-α with an additional 6 histidines (underlined) between the start codon and the first amino acid. Amino acids changed in the TNF-α mutants are shown in bold.

Figure 7 depicts the position and the amino acid changes in the TNF-α mutants.

Figure 8 depicts the % TNF-α activity of the mutants listed in Figure 7. The "oligo name" is based on the changed amino acid in the mutant and the position where the change was generated.

Figures 9A and B depict the reproducibility of the TNF-α activity of the mutants.

Figure 10 depicts the mutation pattern of TNF-α protein sequences. The probability table shows only the amino acid residues of positions 72, 73, 75, 86, 87, 97 and 137. The occurrence of each amino acid residue at a given position is indicated as a relative probability. For example, at position 137, the wild-type amino acid is asparagine; in the TNF-α variants, aspartic is the preferred amino acid at this position.

Figure 11 depicts another example of the mutation pattern of TNF-α protein sequences. The probability table shows only the amino acid residues of positions 21, 30, 31, 32, 33, 35, 65, 66, 67, 111, 112, 115, 140,143, 144, 145, 146 and 147. The occurrence of each amino acid residue at a given position is indicated as a relative probability. For example, at position 21, the wild-type amino acid is glutamine; in the TNF-α variants, arginine is the preferred amino acid at this position.

Figure 12 depicts trimerization domains from TRAF proteins.

Figure 13 depicts the synthesis of a full-length gene and all possible mutations by PCR. Overlapping oligonucleotides corresponding to the full-length gene (black bar, Step 1) and comprising one or more desired mutations are synthesized, heated and annealed. Addition of DNA polymerase to the annealed oligonucleotides results in the 5' to 3' synthesis of DNA (Step 2) to produce longer DNA fragments (Step 3). Repeated cycles of heating, annealing, and DNA synthesis (Step 4) result in the production of longer DNA, including some full-length molecules. These can be selected by a second round of PCR using primers (indicated by arrows) corresponding to the end of the full-length gene (Step 5).

Figure 14 depicts a preferred method for synthesizing a library of the variant TNF-α proteins of the invention using the wild type gene.

Figures 15A and 15B depict an overlapping extension method. At the top of Figure 15A is the template DNA showing the locations of the regions to be mutated (black boxes) and the binding sites of the relevant primers (arrows). The primers R1 and R2 represent a pool of primers, each containing a different mutation; as described herein, this may be done using different ratios of primers if desired. The variant position is flanked by regions of homology sufficient to get hybridization. In this example, three separate PCR reactions are done for step 1. The first reaction contains the template plus oligos

F1 and R1. The second reaction contains template plus F2 and R2, and the third contains the template and F3 and R3. The reaction products are shown. In Step 2, the products from Step 1 tube 1 and Step 1 tube 2 are taken. After purification away from the primers, these are added to a fresh PCR reaction together with F1 and R4. During the denaturation phase of the PCR, the overlapping regions anneal and the second strand is synthesized. T he product is then amplified by the outside primers. In Step 3, the purified product from Step 2 is used in a third PCR reaction, together with the product of Step 1, tube 3 and the primers F1 and R3. The final product corresponds to the full length gene and contains the required mutations.

Figures 16A and 16B depict a ligation of PCR reaction products to synthesize the libraries of the invention. In this technique, the primers also contain an endonuclease restriction site (RE), either blunt, 5' overhanging or 3' overhanging. We set up three separate PCR reactions for Step 1. The first reaction contains the template plus oligos F1 and R1. The second reaction contains template plus F2 and R2, and the third contains the template and F3 and R3. The reaction products are shown. In Step 2, the products of step 1 are purified and then digested with the appropriate restriction endonuclease. The digestion products from Step 2, tube 1 and Step 2, tube 2 and ligate them togther with DNA ligase (step 3). The products are then amplified in Step 4 using primer F1 and R4. The whole process is then repeated by digesting the amplified products, ligating them to the digested products of Step 2, tube 3, and amplifying the final product by primers F1 and R3. It would also be possible to ligate all three PCR products from Step 1 together in one reaction, providing the two restriction sites (RET and RE2) were different.

Figure 17 depicts blunt end ligation of PCR products. In this technique, the primers such as F1 and R1 do not overlap, but they abut. Again three separate PCR reactions are performed. The products from tube 1 and tube 2 are ligated, and then amplified with outside primers F1 and R4. This product is then .l gated with the product from Step 1, tube 3. The final products are then amplified with primers F1 and R3.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to novel proteins and nucleic acids possessing TNF-α antagonist activity. The proteins are generated using a system previously described in WO98/47089 and U.S.S.Nos. 09/058,459, 09/127,926, 60/104,612, 60/158,700, 09/419,351, 60/181,630, 60/186,904, 09/419,351, and an application entitled "Protein Design Automation for Protein Libraries", filed February 12, 2001 (no U.S. serial number received yet) all of which are expressly incorporated by reference in their entirety. In general, these applications describe a variety of computational modeling systems that allow the generation of extremely stable proteins. In this way, variants of TNF proteins

are generated that act as antagonists for wild-type TNF-α. Variant TNF-proteins may be generated from wild-type TNF-α, p55 TNF-R and p75 TNF-R proteins, with preferred embodiments including variant TNF-α proteins.

Generally, there are a variety of computational methods that can be used to generate the variant TNF proteins of the invention. In a preferred embodiment, sequence based methods are used. Alternatively, structure based methods, such as PDA, described in detail below, are used.

Similarly, molecular dynamics calculations can be used to computationally screen sequences by individually calculating mutant sequence scores and compiling a rank ordered list.

In a preferred embodiment, residue pair potentials can be used to score sequences (Miyazawa et al., Macromolecules 18(3):534-552 (1985), expressly incorporated by reference) during computational screening.

In a preferred embodiment, sequence profile scores (Bowie et al., Science 253(5016):164-70 (1991), incorporated by reference) and/or potentials of mean force (Hendlich et al., J. Mol. Biol. 216(1):167-180 (1990), also incorporated by reference) can also be calculated to score sequences. These methods assess the match between a sequence and a 3D protein structure and hence can act to screen for fidelity to the protein structure. By using different scoring functions to rank sequences, different regions of sequence space can be sampled in the computational screen.

Furthermore, scoring functions can be used to screen for sequences that would create metal or co-factor binding sites in the protein (Hellinga, Fold Des. 3(1):R1-8 (1998), hereby expressly incorporated by reference). Similarly, scoring functions can be used to screen for sequences that would create disulfide bonds in the protein. These potentials attempt to specifically modify a protein structure to introduce a new structural motif.

In a preferred embodiment, sequence and/or structural alignment programs can be used to generate the variant TNF-α proteins of the invention. As is known in the art, there are a number of sequence-based alignment programs; including for example, Smith-Waterman searches, Needleman-Wunsch, Double Affine Smith-Waterman, frame search, Gribskov/GCG profile search, Gribskov/GCG profile scan, profile frame search, Bucher generalized profiles, Hidden Markov models, Hframe, Double Frame, Blast, Psi-Blast, Clustal, and GeneWise.

As is known in the art, there are a number of sequence alignment methodologies that can be used. For example, sequence homology based alignment methods can be used to create sequence

alignments of proteins related to the target structure (Altschul et al., J. Mol. Biol. 215(3):403-410 (1990), Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997), both incorporated by reference). These sequence alignments are then examined to determine the observed sequence variations. These sequence variations are tabulated to define a set of variant TNF-α proteins.

Sequence based alignments can be used in a variety of ways. For example, a number of related proteins can be aligned, as is known in the art, and the "variable" and "conserved" residues defined; that is, the residues that vary or remain identical between the family members can be defined. These results can be used to generate a probability table, as outlined below. Similarly, these sequence variations can be tabulated and a secondary library defined from them as defined below. Alternatively, the allowed sequence variations can be used to define the amino acids considered at each position during the computational screening. Another variation is to bias the score for amino acids that occur in the sequence alignment, thereby increasing the likelihood that they are found during computational screening but still allowing consideration of other amino acids. This bias would result in a focused library of variant TNF-α proteins but would not eliminate from consideration amino acids not found in the alignment. In addition, a number of other types of bias may be introduced. For example, diversity may be forced; that is, a "conserved" residue is chosen and altered to force diversity on the protein and thus sample a greater portion of the sequence space. Alternatively, the positions of high variability between family members (i.e. low conservation) can be randomized, either using all or a subset of amino acids. Similarly, outlier residues, either positional outliers or side chain outliers, may be eliminated.

Similarly, structural alignment of structurally related proteins can be done to generate sequence alignments (Orengo et al., Structure 5(8):1093-108 (1997); Holm et al., Nucleic Acids Res. 26(1):316-9 (1998), both of which are incorporated by reference). These sequence alignments can then be examined to determine the observed sequence variations. Libraries can be generated by predicting secondary structure from sequence, and then selecting sequences that are compatible with the predicted secondary structure. There are a number of secondary structure prediction methods such as helix-coil transition theory (Munoz and Serrano, Biopolymers 41:495, 1997), neural networks, local structure alignment and others (e.g., see in Selbig et al., Bioinformatics 15:1039-46, 1999).

Similarly, as outlined above, other computational methods are known, including, but not limited to, sequence profiling [Bowie and Eisenberg, Science 253(5016):164-70, (1991)], rotamer library selections [Dahiyat and Mayo, Protein Sci. 5(5):895-903 (1996); Dahiyat and Mayo, Science 278(5335):82-7 (1997); Desjarlais and Handel, Protein Science 4:2006-2018 (1995); Harbury et al, Proc. Natl. Acad. Sci. U.S.A. 92(18):8408-8412 (1995); Kono et al., Proteins: Structure, Function and Genetics 19:244-255 (1994); Hellinga and Richards, Proc. Natl. Acad. Sci. U.S.A. 91:5803-5807

(1994)]; and residue pair potentials [Jones, Protein Science 3: 567-574, (1994)]; PROSA [Heindlich et al., J. Mol. Biol. 216:167-180 (1990)]; THREADER [Jones et al., Nature 358:86-89 (1992)], and other inverse folding methods such as those described by Simons et al. [Proteins, 34:535-543, (1999)], Levitt and Gerstein [Proc. Natl. Acad. Sci. U.S.A., 95:5913-5920, (1998)], Godzik and Skolnick [Proc. Natl. Acad. Sci. U.S.A., 89:12098-102, (1992)], Godzik et al. [J. Mol. Biol. 227:227-38, (1992)] and two profile methods [Gribskov et al. Proc. Natl. Acad. Sci. U.S.A. 84:4355-4358 (1987) and Fischer and Eisenberg, Protein Sci. 5:947-955 (1996), Rice and Eisenberg J. Mol. Biol. 267:1026-1038(1997)], all of which are expressly incorporated by reference. In addition, other computational methods such as those described by Koehl and Levitt (J. Mol. Biol. 293:1161-1181 (1999); J. Mol. Biol. 293:1183-1193 (1999); expressly incorporated by reference) can be used to create a variant TNF-α library which can optionally then be used to generate a smaller secondary library for use in experimental screening for improved properties and function. In addition, there are computational methods based on forcefield calculations such as SCMF that can be used as well for SCMF, see Delarue et al. Pac. Symp. Biocomput. 109-21 (1997); Koehl et al., J. Mol. Biol. 239:249-75 (1994); Koehl et al., Nat. Struct. Biol. 2:163-70 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222-6 (1996); Koehl et al., J. Mol. Biol. 293:1183-93 (1999); Koehl et al., J. Mol. Biol. 293:1161-81 (1999); Lee J., Mol. Biol. 236:918-39 (1994); and Vasquez Biopolymers 36:53-70 (1995); all of which are expressly incorporated by reference. Other forcefield calculations that can be used to optimize the conformation of a sequence within a computational method, or to generate de novo optimized sequences as outlined herein include, but are not limited to, OPLS-AA [Jorgensen et al., J. Am. Chem. Soc. 118:11225-11236 (1996); Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)]; OPLS [Jorgensen et al., J. Am. Chem. Soc.110:1657ff (1988); Jorgensen et al., J Am. Chem. Soc.112:4768ff (1990)]; UNRES (United Residue Forcefield; Liwo et al., Protein Science 2:1697-1714 (1993); Liwo et al., Protein Science 2:1715-1731 (1993); Liwo et al., J. Comp. Chem. 18:849-873 (1997); Liwo et al., J. Comp. Chem. 18:874-884 (1997); Liwo et al., J. Comp. Chem. 19:259-276 (1998); Forcefield for Protein Structure Prediction (Liwo et al., Proc. Natl. Acad. Sci. U.S.A . 96:5482-5485 (1999)]; ECEPP/3 [Liwo et al., J Protein Chem. 13(4):375-80 (1994)]; AMBER 1.1 force field (Weiner et al., J. Am. Chem. Soc. 106:765-784); AMBER 3.0 force field [U.C. Singh et al., Proc. Natl. Acad. Sci. U.S.A.. 82:755-759 (1985)]; CHARMM and CHARMM22 (Brooks et al., J. Comp. Chem. 4:187-217); cvff3.0 [Dauber-Osguthorpe et al., Proteins: Structure, Function and Genetics, 4:31-47 (1988)]; cff91 (Maple et al., J. Comp. Chem. 15:162-182); also, the DISCOVER (cvff and cff91) and AMBER forcefields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARMM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California), all of which are expressly incorporated by reference. In fact, as is outlined below, these forcefield methods may be used to generate the variant TNF-α library directly; these methods can be used to generate a probability table from which an additional library is directly generated.

In a preferred embodiment, the computational method used to generate the set or library of variant TNF-α proteins is Protein Design Automation (PDA), as is described in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, 60/104,612, 60/158,700, 09/419,351, 60/181630, 60/186,904, 09/419,351, and an application entitled "Protein Design Automation for Protein Libraries", filed February 12, 2001 (no U.S. serial number received yet) and PCT US98/07254, all of which are expressly incorporated herein by reference. Briefly, PDA can be described as follows. A known protein structure is used as the starting point. The residues to be optimized are then identified, which may be the entire sequence or subset(s) thereof. The side chains of any positions to be varied are then removed. The resulting structure consisting of the protein backbone and the remaining sidechains is called the template. Each variable residue position is then preferably classified as a core residue, a surface residue, or a boundary residue; each classification defines a subset of possible amino acid residues for the position (for example, core residues generally will be selected from the set of hydrophobic residues, surface residues generally will be selected from the hydrophilic residues, and boundary residues may be either). Each amino acid can be represented by a discrete set of all allowed conformers of each side chain, called rotamers. Thus, to arrive at an optimal sequence for a backbone, all possible sequences of rotamers must be screened, where each backbone position can be occupied either by each amino acid in all its possible rotameric states, or a subset of amino acids, and thus a subset of rotamers.

Two sets of interactions are then calculated for each rotamer at every position: the interaction of the rotamer side chain with all or part of the backbone (the "singles" energy, also called the rotamer/template or rotamer/backbone energy), and the interaction of the rotamer side chain with all other possible rotamers at every other position or a subset of the other positions (the "doubles" energy, also called the rotamer/rotamer energy). The energy of each of these interactions is calculated through the use of a variety of scoring functions, which include the energy of van der Waal's forces, the energy of hydrogen bonding, the energy of secondary structure propensity, the energy of surface area solvation and the electrostatics. Thus, the total energy of each rotamer interaction, both with the backbone and other rotamers, is calculated, and stored in a matrix form.

The discrete nature of rotamer sets allows a simple calculation of the number of rotamer sequences to be tested. A backbone of length n with m possible rotamers per position will have m^n possible rotamer sequences, a number which grows exponentially with sequence length and renders the calculations either unwieldy or impossible in real time. Accordingly, to solve this combinatorial search problem, a "Dead End Elimination" (DEE) calculation is performed. The DEE calculation is based on the fact that if the worst total interaction of a first rotamer is still better than the best total interaction of a second rotamer, then the second rotamer cannot be part of the global optimum solution. Since the energies of all rotamers have already been calculated, the DEE approach only requires sums over the sequence

9

length to test and eliminate rotamers, which speeds up the calculations considerably. DEE can be rerun comparing pairs of rotamers, or combinations of rotamers, which will eventually result in the determination of a single sequence which represents the global optimum energy.

Once the global solution has been found, a Monte Carlo search may be done to generate a rank-ordered list of sequences in the neighborhood of the DEE solution. Starting at the DEE solution, random positions are changed to other rotamers, and the new sequence energy is calculated. If the new sequence meets the criteria for acceptance, it is used as a starting point for another jump. After a predetermined number of jumps, a rank-ordered list of sequences is generated. Monte Carlo searching is a sampling technique to explore sequence space around the global minimum or to find new local minima distant in sequence space. As is more additionally outlined below, there are other sampling techniques that can be used, including Boltzman sampling, genetic algorithm techniques and simulated annealing. In addition, for all the sampling techniques, the kinds of jumps allowed can be altered (e.g. random jumps to random residues, biased jumps (to or away from wild-type, for example), jumps to biased residues (to or away from similar residues, for example), etc.). Similarly, for all the sampling techniques, the acceptance criteria of whether a sampling jump is accepted can be altered.

As outlined in U.S.S.N. 09/127,926, the protein backbone (comprising (for a naturally occurring protein) the nitrogen, the carbonyl carbon, the α-carbon, and the carbonyl oxygen, along with the direction of the vector from the α-carbon to the β-carbon) may be altered prior to the computational analysis, by varying a set of parameters called supersecondary structure parameters.

Once a protein structure backbone is generated (with alterations, as outlined above) and input into the computer, explicit hydrogens are added if not included within the structure (for example, if the structure was generated by X-ray crystallography, hydrogens must be added). After hydrogen addition, energy minimization of the structure is run, to relax the hydrogens as well as the other atoms, bond angles and bond lengths. In a preferred embodiment, this is done by doing a number of steps of conjugate gradient minimization [Mayo et al., J. Phys. Chem. 94:8897 (1990)] of atomic coordinate positions to minimize the Dreiding force field with no electrostatics. Generally from about 10 to about 250 steps is preferred, with about 50 being most preferred.

The protein backbone structure contains at least one variable residue position. As is known in the art, the residues, or amino acids, of proteins are generally sequentially numbered starting with the N-terminus of the protein. Thus a protein having a methionine at it's N-terminus is said to have a methionine at residue or amino acid position 1, with the next residues as 2, 3, 4, etc. At each position, the wild type (i.e. naturally occurring) protein may have one of at least 20 amino acids, in any number of rotamers. By "variable residue position" herein is meant an amino acid position of the protein to be

designed that is not fixed in the design method as a specific residue or rotamer, generally the wild-type residue or rotamer.

In a preferred embodiment, all of the residue positions of the protein are variable. That is, every amino acid side chain may be altered in the methods of the present invention. This is particularly desirable for smaller proteins, although the present methods allow the design of larger proteins as well. While there is no theoretical limit to the length of the protein which may be designed this way, there is a practical computational limit.

In an alternate preferred embodiment, only some of the residue positions of the protein are variable, and the remainder are "fixed", that is, they are identified in the three dimensional structure as being in a set conformation. In some embodiments, a fixed position is left in its original conformation (which may or may not correlate to a specific rotamer of the rotamer library being used). Alternatively, residues may be fixed as a non-wild type residue; for example, when known site-directed mutagenesis techniques have shown that a particular residue is desirable (for example, to eliminate a proteolytic site or alter the substrate specificity of an enzyme), the residue may be fixed as a particular amino acid. Alternatively, the methods of the present invention may be used to evaluate mutations de novo, as is discussed below. In an alternate preferred embodiment, a fixed position may be "floated"; the amino acid at that position is fixed, but different rotamers of that amino acid are tested. In this embodiment, the variable residues may be at least one, or anywhere from 0.1% to 99.9% of the total number of residues. Thus, for example, it may be possible to change only a few (or one) residues, or most of the residues, with all possibilities in between.

In a preferred embodiment, residues which can be fixed include, but are not limited to, structurally or biologically functional residues; alternatively, biologically functional residues may specifically not be fixed. For example, residues which are known to be important for biological activity, such as the residues which the binding site for a binding partner (ligand/receptor, antigen/antibody, etc.), phosphorylation or glycosylation sites which are crucial to biological function, or structurally important residues, such as disulfide bridges, metal binding sites, critical hydrogen bonding residues, residues critical for backbone conformation such as proline or glycine, residues critical for packing interactions, etc. may all be fixed in a conformation or as a single rotamer, or "floated".

Similarly, residues which may be chosen as variable residues may be those that confer undesirable biological attributes, such as susceptibility to proteolytic degradation, dimerization or aggregation sites, glycosylation sites which may lead to immune responses, unwanted binding activity, unwanted allostery, undesirable enzyme activity but with a preservation of binding, etc. In the present invention, it is the tetramerization domain residues which are varied, as outlined below.

11

In a preferred embodiment, each variable position is classified as either a core, surface or boundary residue position, although in some cases, as explained below, the variable position may be set to glycine to minimize backbone strain. In addition, as outlined herein, residues need not be classified, they can be chosen as variable and any set of amino acids may be used. Any combination of core, surface and boundary positions can be utilized: core, surface and boundary residues; core and surface residues; core and boundary residues, and surface and boundary residues, as well as core residues alone, surface residues alone, or boundary residues alone.

The classification of residue positions as core, surface or boundary may be done in several ways, as will be appreciated by those in the art. In a preferred embodiment, the classification is done via a visual scan of the original protein backbone structure, including the side chains, and assigning a classification based on a subjective evaluation of one skilled in the art of protein modeling. Alternatively, a preferred embodiment utilizes an assessment of the orientation of the $C\alpha$-$C\beta$ vectors relative to a solvent accessible surface computed using only the template $C\alpha$ atoms, as outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926 60/104,612, 60/158,700, 09/419,351, 60/181630, 60/186,904, 09/419,351and an application entitled "Protein Design Automation for Protein Libraries" filed February 12, 2001 (no U.S. serial number received yet) and PCT US98/07254. Alternatively, a surface area calculation can be done.

Once each variable position is classified as either core, surface or boundary, a set of amino acid side chains, and thus a set of rotamers, is assigned to each position. That is, the set of possible amino acid side chains that the program will allow to be considered at any particular position is chosen. Subsequently, once the possible amino acid side chains are chosen, the set of rotamers that will be evaluated at a particular position can be determined. Thus, a core residue will generally be selected from the group of hydrophobic residues consisting of alanine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine (in some embodiments, when the α scaling factor of the van der Waals scoring function, described below, is low, methionine is removed from the set), and the rotamer set for each core position potentially includes rotamers for these eight amino acid side chains (all the rotamers if a backbone independent library is used, and subsets if a rotamer dependent backbone is used). Similarly, surface positions are generally selected from the group of hydrophilic residues consisting of alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine and histidine. The rotamer set for each surface position thus includes rotamers for these ten residues. Finally, boundary positions are generally chosen from alanine, serine, threonine, aspartic acid, asparagine, glutamine, glutamic acid, arginine, lysine histidine, valine, isoleucine, leucine, phenylalanine, tyrosine, tryptophan, and methionine. The rotamer set for each boundary position thus potentially includes every rotamer for these seventeen residues (assuming cysteine, glycine and proline are not used, although they can be). Additionally, in some preferred embodiments,

12

a set of 18 naturally occurring amino acids (all except cysteine and proline, which are known to be particularly disruptive) are used.

Thus, as will be appreciated by those in the art, there is a computational benefit to classifying the residue positions, as it decreases the number of calculations. It should also be noted that there may be situations where the sets of core, boundary and surface residues are altered from those described above; for example, under some circumstances, one or more amino acids is either added or subtracted from the set of allowed amino acids. For example, some proteins which dimerize or multimerize, or have ligand binding sites, may contain hydrophobic surface residues, etc. In addition, residues that do not allow helix "capping" or the favorable interaction with an α-helix dipole may be subtracted from a set of allowed residues. This modification of amino acid groups is done on a residue by residue basis.

In a preferred embodiment, proline, cysteine and glycine are not included in the list of possible amino acid side chains, and thus the rotamers for these side chains are not used. However, in a preferred embodiment, when the variable residue position has a ϕ angle (that is, the dihedral angle defined by 1) the carbonyl carbon of the preceding amino acid; 2) the nitrogen atom of the current residue; 3) the α-carbon of the current residue; and 4) the carbonyl carbon of the current residue) greater than 0°, the position is set to glycine to minimize backbone strain.

Once the group of potential rotamers is assigned for each variable residue position, processing proceeds as outlined in U.S.S.N. 09/127,926 and PCT US98/07254. This processing step entails analyzing interactions of the rotamers with each other and with the protein backbone to generate optimized protein sequences. Simplistically, the processing initially comprises the use of a number of scoring functions to calculate energies of interactions of the rotamers, either to the backbone itself or other rotamers. Preferred PDA scoring functions include, but are not limited to, a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, a secondary structure propensity scoring function and an electrostatic scoring function. As is further described below, at least one scoring function is used to score each position, although the scoring functions may differ depending on the position classification or other considerations, like favorable interaction with an α-helix dipole. As outlined below, the total energy which is used in the calculations is the sum of the energy of each scoring function used at a particular position, as is generally shown in Equation 1:

<div align="center">

Equation 1

$$E_{total} = nE_{vdw} + nE_{as} + nE_{h\text{-}bonding} + nE_{ss} + nE_{elec}$$

</div>

In Equation 1, the total energy is the sum of the energy of the van der Waals potential (E_{vdw}), the energy of atomic solvation (E_{as}), the energy of hydrogen bonding ($E_{h\text{-}bonding}$), the energy of secondary structure (E_{ss}) and the energy of electrostatic interaction (E_{elec}). The term n is either 0 or 1, depending on whether the term is to be considered for the particular residue position.

As outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, 60/104,612, 60/158,700, 09/419,351, 60/181630, 60/186,904, 09/419,351, and an application entitled "Protein Design Automation for Protein Libraries", filed February 12, 2001 (no U.S. serial number received yet) and PCT US98/07254, any combination of these scoring functions, either alone or in combination, may be used. Once the scoring functions to be used are identified for each variable position, the preferred first step in the computational analysis comprises the determination of the interaction of each possible rotamer with all or part of the remainder of the protein. That is, the energy of interaction, as measured by one or more of the scoring functions, of each possible rotamer at each variable residue position with either the backbone or other rotamers, is calculated. In a preferred embodiment, the interaction of each rotamer with the entire remainder of the protein, i.e. both the entire template and all other rotamers, is done. However, as outlined above, it is possible to only model a portion of a protein, for example a domain of a larger protein, and thus in some cases, not all of the protein need be considered. The term "portion", or similar grammatical equivalents thereof, as used herein, with regard to a protein refers to a fragment of that protein. This fragment may range in size from 6-10 amino acid residues to the entire amino acid sequence minus one amino acid. Accordingly, the term "portion", as used herein, with regard to a nucleic refers to a fragment of that nucleic acid. This fragment may range in size from 10 nucleotides to the entire nucleic acid sequence minus one nucleotide.

In a preferred embodiment, the first step of the computational processing is done by calculating two sets of interactions for each rotamer at every position: the interaction of the rotamer side chain with the template or backbone (the "singles" energy), and the interaction of the rotamer side chain with all other possible rotamers at every other position (the "doubles" energy), whether that position is varied or floated. It should be understood that the backbone in this case includes both the atoms of the protein structure backbone, as well as the atoms of any fixed residues, wherein the fixed residues are defined as a particular conformation of an amino acid.

Thus, "singles" (rotamer/template) energies are calculated for the interaction of every possible rotamer at every variable residue position with the backbone, using some or all of the scoring functions. Thus, for the hydrogen bonding scoring function, every hydrogen bonding atom of the rotamer and every hydrogen bonding atom of the backbone is evaluated, and the E_{HB} is calculated for each possible rotamer at every variable position. Similarly, for the van der Waals scoring function, every atom of the

14

rotamer is compared to every atom of the template (generally excluding the backbone atoms of its own residue), and the E_{vdw} is calculated for each possible rotamer at every variable residue position. In addition, generally no van der Waals energy is calculated if the atoms are connected by three bonds or less. For the atomic solvation scoring function, the surface of the rotamer is measured against the surface of the template, and the E_{as} for each possible rotamer at every variable residue position is calculated. The secondary structure propensity scoring function is also considered as a singles energy, and thus the total singles energy may contain an E_{ss} term. As will be appreciated by those in the art, many of these energy terms will be close to zero, depending on the physical distance between the rotamer and the template position; that is, the farther apart the two moieties, the lower the energy.

For the calculation of "doubles" energy (rotamer/rotamer), the interaction energy of each possible rotamer is compared with every possible rotamer at all other variable residue positions. Thus, "doubles" energies are calculated for the interaction of every possible rotamer at every variable residue position with every possible rotamer at every other variable residue position, using some or all of the scoring functions. Thus, for the hydrogen bonding scoring function, every hydrogen bonding atom of the first rotamer and every hydrogen bonding atom of every possible second rotamer is evaluated, and the E_{HB} is calculated for each possible rotamer pair for any two variable positions. Similarly, for the van der Waals scoring function, every atom of the first rotamer is compared to every atom of every possible second rotamer, and the E_{vdw} is calculated for each possible rotamer pair at every two variable residue positions. For the atomic solvation scoring function, the surface of the first rotamer is measured against the surface of every possible second rotamer, and the E_{as} for each possible rotamer pair at every two variable residue positions is calculated. The secondary structure propensity scoring function need not be run as a "doubles" energy, as it is considered as a component of the "singles" energy. As will be appreciated by those in the art, many of these double energy terms will be close to zero, depending on the physical distance between the first rotamer and the second rotamer; that is, the farther apart the two moieties, the lower the energy.

In addition, as will be appreciated by those in the art, a variety of force fields that can be used in the PDA calculations can be used, including, but not limited to, Dreiding I and Dreiding II [Mayo et al, J. Phys. Chem. 94:8897 (1990)], AMBER [Weiner et al., J. Amer. Chem. Soc. 106:765 (1984) and Weiner et al., J. Comp. Chem. 106:230 (1986)], MM2 [Allinger, J. Chem. Soc. 99:8127 (1977), Liljefors et al., J. Com. Chem. 8:1051 (1987)]; MMP2 [Sprague et al., J. Comp. Chem. 8:581 (1987)]; CHARMM [Brooks et al., J. Comp. Chem. 106:187 (1983)]; GROMOS; and MM3 [Allinger et al., J. Amer. Chem. Soc. 111:8551 (1989)], OPLS-AA [Jorgensen et al., J. Am. Chem. Soc. 118:11225-11236 (1996); Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)]; OPLS [Jorgensen et al., J. Am. Chem. Soc.110:1657ff (1988); Jorgensen et al., J Am. Chem. Soc. 112:4768ff (1990)]; UNRES (United Residue Forcefield; Liwo et al., Protein Science

15

2:1697-1714 (1993); Liwo et al., Protein Science 2:1715-1731 (1993); Liwo et al., J. Comp. Chem. 18:849-873 (1997); Liwo et al., J. Comp. Chem. 18:874-884 (1997); Liwo et al., J. Comp. Chem. 19:259-276 (1998); Forcefield for Protein Structure Prediction (Liwo et al., Proc. Natl. Acad. Sci. U.S.A 96:5482-5485 (1999)]; ECEPP/3 [Liwo et al., J Protein Chem. 13(4):375-80 (1994)]; AMBER 1.1 force field (Weiner, et al., J. Am. Chem. Soc. 106:765-784); AMBER 3.0 force field (U.C. Singh et al., Proc. Natl. Acad. Sci. U.S.A.. 82:755-759); CHARMM and CHARMM22 (Brooks et al., J. Comp. Chem. 4:187-217); cvff3.0 [Dauber-Osguthorpe, et al., Proteins: Structure, Function and Genetics, 4:31-47 (1988)]; cff91 (Maple, et al., J. Comp. Chem. 15:162-182); also, the DISCOVER (cvff and cff91) and AMBER forcefields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARMM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California), all of which are expressly incorporated by reference.

Once the singles and doubles energies are calculated and stored, the next step of the computational processing may occur. As outlined in U.S.S.N. 09/127,926 and PCT US98/07254, preferred embodiments utilize a Dead End Elimination (DEE) step, and preferably a Monte Carlo step.

PDA, viewed broadly, has three components that may be varied to alter the output (e.g. the primary library): the scoring functions used in the process; the filtering technique, and the sampling technique.

In a preferred embodiment, the scoring functions may be altered. In a preferred embodiment, the scoring functions outlined above may be biased or weighted in a variety of ways. For example, a bias towards or away from a reference sequence or family of sequences can be done; for example, a bias towards wild-type or homolog residues may be used. Similarly, the entire protein or a fragment of it may be biased; for example, the active site may be biased towards wild-type residues, or domain residues towards a particular desired physical property can be done. Furthermore, a bias towards or against increased energy can be generated. Additional scoring function biases include, but are not limited to applying electrostatic potential gradients or hydrophobicity gradients, adding a substrate or binding partner to the calculation, or biasing towards a desired charge or hydrophobicity.

In addition, in an alternative embodiment, there are a variety of additional scoring functions that may be used. Additional scoring functions include, but are not limited to torsional potentials, or residue pair potentials, or residue entropy potentials. Such additional scoring functions can be used alone, or as functions for processing the library after it is scored initially. For example, a variety of functions derived from data on binding of peptides to MHC (Major Histocompatibility Complex) can be used to rescore a library in order to eliminate proteins containing sequences which can potentially bind to MHC, i.e. potentially immunogenic sequences.

In a preferred embodiment, a variety of filtering techniques can be done, including, but not limited to, DEE and its related counterparts. Additional filtering techniques include, but are not limited to branch-and-bound techniques for finding optimal sequences (Gordon and Mayo, Structure Fold. Des. 7:1089-98, 1999), and exhaustive enumeration of sequences.

As will be appreciated by those in the art, once an optimized sequence or set of sequences is generated, a variety of sequence space sampling methods can be done, either in addition to the preferred Monte Carlo methods, or instead of a Monte Carlo search. That is, once a sequence or set of sequences is generated, preferred methods utilize sampling techniques to allow the generation of additional, related sequences for testing.

These sampling methods can include the use of amino acid substitutions, insertions or deletions, or recombinations of one or more sequences. As outlined herein, a preferred embodiment utilizes a Monte Carlo search, which is a series of biased, systematic, or random jumps. However, there are other sampling techniques that can be used, including Boltzman sampling, genetic algorithm techniques and simulated annealing. In addition, for all the sampling techniques, the kinds of jumps allowed can be altered (e.g. random jumps to random residues, biased jumps (to or away from wild-type, for example), jumps to biased residues (to or away from similar residues, for example, etc.). Jumps where multiple residue positions are coupled (two residues always change together, or never change together), jumps where whole sets of residues change to other sequences (e.g., recombination). Similarly, for all the sampling techniques, the acceptance criteria of whether a sampling jump is accepted can be altered.

In addition, it should be noted that the preferred methods of the invention result in a rank ordered list of sequences; that is, the sequences are ranked on the basis of some objective criteria. However, as outlined herein, it is possible to create a set of non-ordered sequences, for example by generating a probability table directly (for example using SCMF analysis or sequence alignment techniques) that lists sequences without ranking them. The sampling techniques outlined herein can be used in either situation.

In a preferred embodiment, Boltzman sampling is done. As will be appreciated by those in the art, the temperature criteria for Boltzman sampling can be altered to allow broad searches at high temperature and narrow searches close to local optima at low temperatures (see e.g., Metropolis et al., J. Chem. Phys. 21:1087, 1953).

In a preferred embodiment, the sampling technique utilizes genetic algorithms, e.g., such as those described by Holland (Adaptation in Natural and Artificial Systems, 1975, Ann Arbor, U. Michigan

Press). Genetic algorithm analysis generally takes generated sequences and recombines them computationally, similar to a nucleic acid recombination event, in a manner similar to "gene shuffling". Thus the "jumps" of genetic algorithm analysis generally are multiple position jumps. In addition, as outlined below, correlated multiple jumps may also be done. Such jumps can occur with different crossover positions and more than one recombination at a time, and can involve recombination of two or more sequences. Furthermore, deletions or insertions (random or biased) can be done. In addition, as outlined below, genetic algorithm analysis may also be used after the secondary library has been generated.

In a preferred embodiment, the sampling technique utilizes simulated annealing, e.g., such as described by Kirkpatrick et al. [Science, 220:671-680 (1983)]. Simulated annealing alters the cutoff for accepting good or bad jumps by altering the temperature. That is, the stringency of the cutoff is altered by altering the temperature. This allows broad searches at high temperature to new areas of sequence space, altering with narrow searches at low temperature to explore regions in detail.

In addition, as outlined below, these sampling methods can be used to further process a first set to generate additional sets of variant TNF-α proteins.

As used herein variant TNF-α proteins include TNF-α monomers.

The computational processing results in a set of optimized variant TNF protein sequences. Optimized variant TNF-α protein sequences are generally different from the wild-type TNF-α sequence in structural regions critical for receptor affinity, e.g. p55, p75. Preferably, each optimized variant TNF-α protein sequence comprises at least about 1 variant amino acid from the starting or wild type sequence, with 3-5 being preferred.

Thus, in the broadest sense, the present invention is directed to variant TNF-α proteins that are antagonists of wild-type TNF-α. By "variant TNF-α proteins" herein is meant TNF-α proteins, which have been designed using the computational methods outlined herein to differ from the corresponding wild-type protein by at least 1 amino acid.

By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures, i.e., "analogs" such as peptoids [see Simon et al., Proc. Natl. Acd. Sci. U.S.A. 89(20:9367-71 (1992)], generally depending on the method of synthesis. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline, and noreleucine are

18

considered amino acids for the purposes of the invention. "Amino acid" also includes imino acid
residues such as proline and hydroxyproline. In addition, any amino acid representing a component of
the variant TNF-α proteins can be replaced by the same amino acid but of the opposite chirality.
Thus, any amino acid naturally occurring in the L-configuration (which may also be referred to as the R
or S, depending upon the structure of the chemical entity) may be replaced with an amino acid of the
same chemical structural type, but of the opposite chirality, generally referred to as the D- amino acid
but which can additionally be referred to as the R- or the S-, depending upon its composition and
chemical configuration. Such derivatives have the property of greatly increased stability, and therefore
are advantageous in the formulation of compounds which may have longer in vivo half lives, when
administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other
routes.   In the preferred embodiment, the amino acids are in the (S) or L-configuration.   If non-
naturally occurring side chains are used, non-amino acid substituents may be used, for example to
prevent or retard in vivo degradations.   Proteins including non-naturally occurring amino acids may be
synthesized or in some cases, made recombinantly; see van Hest et al., FEBS Lett 428:(1-2) 68-70
May 22 1998 and Tang et al., Abstr. Pap Am. Chem. S218:U138-U138 Part 2 August 22, 1999, both of
which are expressly incorporated by reference herein.

Aromatic amino acids may be replaced with D- or L-naphylalanine, D- or L-Phenylglycine, D- or L-2-
thieneylalanine, D- or L-1-, 2-, 3- or 4-pyreneylalanine, D- or L-3-thieneylalanine, D- or L-(2-pyridinyl)-
alanine, D- or L-(3-pyridinyl)-alanine, D- or L-(2-pyrazinyl)-alanine, D- or L-(4-isopropyl)-phenylglycine,
D-(trifluoromethyl)-phenylglycine, D-(trifluoromethyl)-phenylalanine, D-p-fluorophenylalanine, D- or L-
p-biphenylphenylalanine, D- or L-p-methoxybiphenylphenylalanine, D- or L-2-indole(alkyl)alanines,
and D- or L-alkylainines where alkyl may be substituted or unsubstituted methyl, ethyl, propyl, hexyl,
butyl, pentyl, isopropyl, iso-butyl, sec-isotyl, iso-pentyl, non-acidic amino acids, of C1-C20.

Acidic amino acids can be substituted with non-carboxylate amino acids while maintaining a negative
charge, and derivatives or analogs thereof, such as the non-limiting examples of (phosphono)alanine,
glycine, leucine, isoleucine, threonine, or serine; or sulfated (e.g., -SO.sub.3 H) threonine, serine,
tyrosine.

Other substitutions may include unnatural hyroxylated amino acids may made by combining "alkyl"
with any natural amino acid.   The term "alkyl" as used herein refers to a branched or unbranched
saturated hydrocarbon group of 1 to 24 carbon atoms, such as methyl, ethyl, n-propyl, isoptopyl, n-
butyl, isobutyl, t-butyl, octyl, decyl, tetradecyl, hexadecyl, eicosyl, tetracisyl and the like. Alkyl includes
heteroalkyl, with atoms of nitrogen, oxygen and sulfur. Preferred alkyl groups herein contain 1 to 12
carbon atoms. Basic amino acids may be substituted with alkyl groups at any position of the naturally
occurring amino acids lysine, arginine, ornithine, citrulline, or (guanidino)-acetic acid, or other

(guanidino)alkyl-acetic acids, where "alkyl" is define as above. Nitrile derivatives (e.g., containing the CN-moiety in place of COOH) may also be substituted for asparagine or glutamine, and methionine sulfoxide may be substituted for methionine. Methods of preparation of such peptide derivatives are well known to one skilled in the art.

In addition, any amide linkage in any of the variant TNF-α polypeptides can be replaced by a ketomethylene moiety. Such derivatives are expected to have the property of increased stability to degradation by enzymes, and therefore possess advantages for the formulation of compounds which may have increased in vivo half lives, as administered by oral, intravenous, intramuscular, intraperitoneal, topical, rectal, intraocular, or other routes.

Additional amino acid modifications of amino acids of variant TNF-α polypeptides of to the present invention may include the following: Cysteinyl residues may be reacted with alpha-haloacetates (and corresponding amines), such as 2-chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues may also be derivatized by reaction with compounds such as bromotrifluoroacetone, alpha-bromo-beta-(5-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues may be derivatized by reaction with compounds such as diethylprocarbonate e.g., at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain, and para-bromophenacyl bromide may also be used; e.g., where the reaction is preferably performed in 0.1M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues may be reacted with compounds such as succinic or other carboxylic acid anhydrides. Derivatization with these agents is expected to have the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing alpha-amino-containing residues include compounds such as imidoesters/e.g., as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues may be modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin according to known method steps. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

The specific modification of tyrosyl residues per se is well-known, such as for introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. N-acetylimidizol and tetranitromethane may be used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) may be selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl- (4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore aspartyl and glutamyl residues may be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues may be frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues may be deamidated under mildly acidic conditions. Either form of these residues falls within the scope of the present invention.

The TNF-α proteins may be from any number of organisms, with TNF-α proteins from mammals being particularly preferred. Suitable mammals include, but are not limited to, rodents (rats, mice, hamsters, guinea pigs, etc.), primates, farm animals (including sheep, goats, pigs, cows, horses, etc) and in the most preferred embodiment, from humans (the sequence of which is depicted in Figure 6). As will be appreciated by those in the art, TNF-α proteins based on TNF-α proteins from mammals other than humans may find use in animal models of human disease.

The TNF proteins of the invention are antagonists of wild-type TNF-α. By "antagonists of wild-type TNF-α" herein is meant that the variant TNF-α protein inhibits or significantly decreases the activation of receptor signaling by wild-type TNF-α proteins. In a preferred embodiment, the variant TNF-α protein interacts with the wild-type TNF-α protein such that the complex comprising the variant TNF-α and wild-type TNF-α is incapable of activating TNF receptors, i.e. p55 TNF-R or p75 TNF-R.

In a preferred embodiment, the variant TNF-α protein is a variant TNF-α protein which functions as an antagonist of wild-type TNF-α. Preferably, the variant TNF-α protein preferentially interacts with wild-type TNF-α to form mixed trimers with the wild-type protein such that receptor binding does not occur and/or TNF-α signaling is not initiated.

By mixed trimers herein is meant that monomers of wild-type and variant TNF-α proteins interact to form trimeric TNF-α. Mixed trimers may comprise 1 variant TNF-α protein:2 wild-type TNF-α proteins, 2 variant TNF-α proteins:1 wild-type TNF-α protein. In some embodiments, trimers may be formed comprising only variant TNF-α proteins.

21

The variant TNF-α antagonist proteins of the invention are highly specific for TNF-α antagonism relative to TNF-β antagonism. Additional characteristics include improved stability, pharmacokinetics, and high affinity for wild-type TNF-α. Variants with higher affinity toward wild-type TNF-α may be generated from variants exhibiting TNF-α antagonism as outlined above.

In a preferred embodiment, variant TNF-α proteins exhibit decreased biological activity as compared to wild-type TNF-α, including but not limited to, decreased binding to the receptor, decreased activation and/or ultimately a loss of cytotoxic activity. By "cytotoxic activity" herein refers to the ability of wild-type TNF-α to selectively kill or inhibit cells. Variant TNF-α proteins that exhibit less than 50% biological activity are preferred. More preferred are variant TNF-α proteins that exhibit less than 25%, even more preferred are variant proteins that exhibit less than 15%, and most preferred are variant TNF-α proteins that exhibit less than 10% of a biological activity of wild type TNF-α. Suitable assays include, but are not limited to, TNF α cytotoxicity assays, DNA binding assays; transcription assays (using reporter constructs; see Stavridi, supra); size exclusion chromatography assays and radiolabeling/immuno-precipitation; see Corcoran et al., supra); and stability assays (including the use of circular dichroism (CD) assays and equilibrium studies; see Mateu, supra); all of which are expressly incorporated by reference.

In one embodiment, at least one property critical for binding affinity of the variant TNF-α proteins is altered when compared to the same property of wild-type TNF-α and in particular, variant TNF-α proteins with altered receptor affinity are preferred Particularly preferred are variant TNF-α with altered affinity toward oligomerization to wild-type TNF-α.

Thus, the invention provides variant TNF-α proteins with altered binding affinities such that the variant TNF-α proteins will preferentially oligomerize with wild-type TNF-α, but do not substantially interact with wild-type TNF receptors, i.e., p55, p75. "Preferentially" in this case means that given equal amounts of variant TNF-α monomers and wild-type TNF-α monomers, at least 25% of the resulting trimers are mixed trimers of variant and wild-type TNF-α, with at least about 50% being preferred, and at least about 80-90% being particularly preferred. In other words, it is preferable that the variant TNF-α proteins of the invention have greater affinity for wild-type TNF-α protein as compared to wild-type TNF-α proteins. By "do not substantially interact with TNF receptors" herein is meant that the variant TNF-α proteins will not be able to associate with either the p55 or p75 receptors to activate the receptor and initiate the TNF signaling pathway(s).

As outlined above, the invention provides variant TNF-α nucleic acids encoding variant TNF-α polypeptides. The variant TNF-α polypeptide preferably has at least one property, which is substantially different from the same property of the corresponding naturally occurring TNF

polypeptide. The property of the variant TNF-α polypeptide is the result the PDA analysis of the present invention.

The term "altered property" or grammatical equivalents thereof in the context of a polypeptide, as used herein, further refers to any characteristic or attribute of a polypeptide that can be selected or detected and compared to the corresponding property of a naturally occurring protein. These properties include, but are not limited to cytotoxic activity; oxidative stability, substrate specificity, substrate binding or catalytic activity, thermal stability, alkaline stability, pH activity profile, resistance to proteolytic degradation, kinetic association (K_{on}) and dissociation (K_{off}) rate, protein folding, inducing an immune response, ability to bind to a ligand, ability to bind to a receptor, ability to be secreted, ability to be displayed on the surface of a cell, ability to oligomerize, ability to signal, ability to stimulate cell proliferation, ability to inhibit cell proliferation, ability to induce apoptosis, ability to be modified by phosphorylation or glycosylation, ability to treat disease.

Unless otherwise specified, a substantial change in any of the above-listed properties, when comparing the property of a variant TNF-α polypeptide to the property of a naturally occurring TNF protein is preferably at least a 20%, more preferably, 50%, more preferably at least a 2-fold increase or decrease.

A change in cytotoxic activity is evidenced by at least a 75% or greater decrease in cell death initiated by a variant TNF-α protein as compared to wild-type protein.

A change in binding affinity is evidenced by at least a 5% or greater increase or decrease in binding affinity to wild-type TNF receptor proteins or to wild-type TNF-α.

A change in oxidative stability is evidenced by at least about 20%, more preferably at least 50% increase of activity of a variant TNF-α protein when exposed to various oxidizing conditions as compared to that of wild-type TNF-α.. Oxidative stability is measured by known procedures.

A change in alkaline stability is evidenced by at least about a 5% or greater increase or decrease (preferably increase) in the half life of the activity of a variant TNF-α protein when exposed to increasing or decreasing pH conditions as compared to that of wild-type TNF-α. Generally, alkaline stability is measured by known procedures.

A change in thermal stability is evidenced by at least about a 5% or greater increase or decrease (preferably increase) in the half life of the activity of a variant TNF-α protein when exposed to a

relatively high temperature and neutral pH as compared to that of wild-type TNF-α. Generally, thermal stability is measured by known procedures.

Similarly, variant TNF-α proteins, for example are experimentally tested and validated in *in vivo* and in *in vitro* assays. Suitable assays include, but are not limited to, e.g., examining the binding affinity of variant TNF-α proteins as compared to wild-type TNF-α proteins for naturally occurring TNF-α and TNF receptor proteins such as p55 and p75, and can include quantitative comparisons comparing kinetic and equilibrium binding constants. The kinetic association rate (K_{on}) and dissociation rate (K_{off}), and the equilibrium binding constants (K_d) can be determined using surface plasmon resonance on a BIAcore instrument following the standard procedure in the literature [Pearce et al., Biochemistry 38:81-89 (1999)]. Again, as outlined herein, variant TNF-α proteins that preferentially form mixed trimers with wild-type TNF-α proteins, but do not substantially interact with wild-type receptor proteins are preferred.

In a preferred embodiment, the antigenic profile in the host animal of the variant TNF-α protein is similar, and preferably identical, to the antigenic profile of the host TNF-α; that is, the variant TNF-α protein does not significantly stimulate the host organism (e.g. the patient) to an immune response; that is, any immune response is not clinically relevant and there is no allergic response or neutralization of the protein by an antibody. That is, in a preferred embodiment, the variant TNF-α protein does not contain additional or different epitopes from the TNF-α. By 'epitope" or "determinant" herein is meant a portion of a protein which will generate and/or bind an antibody. Thus, in most instances, no significant amount of antibodies are generated to a variant TNF-α protein. In general, this is accomplished by not significantly altering surface residues, as outlined below nor by adding any amino acid residues on the surface which can become glycosylated, as novel glycosylation can result in an immune response.

The variant TNF-α proteins and nucleic acids of the invention are distinguishable from naturally occurring wild-type TNF-α. By "naturally occurring" or "wild type" or grammatical equivalents, herein is meant an amino acid sequence or a nucleotide sequence that is found in nature and includes allelic variations; that is, an amino acid sequence or a nucleotide sequence that usually has not been intentionally modified. Accordingly, by "non-naturally occurring" or "synthetic" or "recombinant" or grammatical equivalents thereof, herein is meant an amino acid sequence or a nucleotide sequence that is not found in nature; that is, an amino acid sequence or a nucleotide sequence that usually has been intentionally modified. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e., using the in vivo cellular machinery of the host cell rather than in vitro manipulations, however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered

recombinant for the purpose of the invention. Representative amino acid and nucleotide sequences of a naturally occurring human TNF-α are shown in Figure 6. It should be noted that unless otherwise stated, all positional numbering of variant TNF-α proteins and variant TNF-α nucleic acids is based on these sequences. That is, as will be appreciated by those in the art, an alignment of TNF-α proteins and variant TNF-α proteins can be done using standard programs, as is outlined below, with the identification of "equivalent" positions between the two proteins. Thus, the variant TNF-α proteins and nucleic acids of the invention are non-naturally occurring; that is, they do not exist in nature.

Thus, in a preferred embodiment, the variant TNF-α protein has an amino acid sequence that differs from a wild-type TNF-α sequence by at least 1-5% of the residues. That is, the variant TNF-α proteins of the invention are less than about 97-99% identical to a wild-type TNF-α amino acid sequence. Accordingly, a protein is a "variant TNF-α protein" if the overall homology of the protein sequence to the amino acid sequence shown in Figure 6 is preferably less than about 99%, more preferably less than about 98%, even more preferably less than about 97% and mor preferably less than 95%. In some embodiments, the homology will be as low as about 75-80%. Stated differently, based on the human TNF sequence of Figure 6, variant TNF-α proteins have at least about 1 residue that differs from the human TNF-α sequence, with at least about 2, 3, 4, or 5 different residues. Preferred variant TNF-α proteins have 3 to 5 different residues.

Homology in this context means sequence similarity or identity, with identity being preferred. As is known in the art, a number of different programs can be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, Adv. Appl. Math., 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, J. Mol. Biol., 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Natl. Acad. Sci. U.S.A., 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence program described by Devereux et al., Nucl. Acid Res., 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc.

An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the

clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, J. Mol. Evol. 35:351-360 (1987); the method is similar to that described by Higgins & Sharp CABIOS 5:151-153 (1989). Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al., J. Mol. Biol. 215, 403-410, (1990); Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997); and Karlin et al., Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787 (1993). A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al., Methods in Enzymology, 266:460-480 (1996); http://blast.wustl/edu/blast/ README.html]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

An additional useful algorithm is gapped BLAST as reported by Altschul et al., Nucl. Acids Res., 25:3389-3402. Gapped BLAST uses BLOSUM-62 substitution scores; threshold T parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of k a cost of $10+k$; X_u set to 16, and X_g set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).

In a similar manner, "percent (%) nucleic acid sequence identity" with respect to the coding sequence of the polypeptides identified herein is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the coding sequence of the cell cycle protein. A preferred method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the protein encoded by the sequence of Figure 6, it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids in relation to the total number of amino

26

acids. Thus, for example, sequence identity of sequences shorter than that shown in Figure 6, as discussed below, will be determined using the number of amino acids in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including gaps are assigned a value of "0", which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity can be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the "shorter" sequence in the aligned region and multiplying by 100. The "longer" sequence is the one having the most actual residues in the aligned region.

Thus, the variant TNF-α proteins of the present invention may be shorter or longer than the amino acid sequence shown in Figure 6B. Thus, in a preferred embodiment, included within the definition of variant TNF proteins are portions or fragments of the sequences depicted herein. Fragments of variant TNF-α proteins are considered variant TNF-α proteins if a0 they share at least one antigenic epitope; b) have at least the indicated homology; c) and preferably have variant TNF-α biological activity as defined herein.

In a preferred embodiment, as is more fully outlined below, the variant TNF-α proteins include further amino acid variations, as compared to a wild type TNF-α, than those outlined herein. In addition, as outlined herein, any of the variations depicted herein may be combined in any way to form additional novel variant TNF-α proteins.

In addition, variant TNF-α proteins can be made that are longer than those depicted in the figures, for example, by the addition of epitope or purification tags, as outlined herein, the addition of other fusion sequences, etc. For example, the variant TNF-α proteins of the invention may be fused to other therapeutic proteins or to other proteins such as Fc or serum albumin for pharmacokinetic purposes. See for example U.S. Patent No. 5,766,883 and 5,876,969, both of which are expressly incorporated by reference.

In a preferred embodiment, the variant TNF-α proteins comprise residues selected from the following positions 21, 30, 31, 32, 33, 35, 65, 66, 67, 111, 112, 115, 140, 143, 144, 145, 146, and 147.

Also included within the invention are variant TNF-α proteins comprising variable residues in core, surface, and boundary residues.

27

Preferred amino acids for each position, including the human TNF-α residues, are shown in Figure 7. Thus, for example, at position 143, preferred amino acids are Glu, Asn, Gln, Ser, Arg, and Lys; etc.

Preferred changes are as follows: D143E, D143N, D143S, A145R, A145K, A145E, E146K, E146R and A84V. These may be done either individually or in combination, with any combination being possible. However, as outlined herein, preferred embodiments utilize at least 1 to 5, and preferably more, positions in each variant TNF-α protein.

In a preferred embodiment, the variant TNF-α proteins of the invention are human TNF-α conformers. By "conformer" herein is meant a protein that has a protein backbone 3D structure that is virtually the same but has significant differences in the amino acid side chains. That is, the variant TNF-α proteins of the invention define a conformer set, wherein all of the proteins of the set share a backbone structure and yet have sequences that differ by at least 1-3-5%. The three dimensional backbone structure of a variant TNF-α protein thus substantially corresponds to the three dimensional backbone structure of human TNF-α. "Backbone" in this context means the non-side chain atoms: the nitrogen, carbonyl carbon and oxygen, and the α-carbon, and the hydrogens attached to the nitrogen and α-carbon. To be considered a conformer, a protein must have backbone atoms that are no more than 2 Å from the human TNF-α structure, with no more than 1.5 Å being preferred, and no more than 1 Å being particularly preferred. In general, these distances may be determined in two ways. In one embodiment, each potential conformer is crystallized and its three dimensional structure determined. Alternatively, as the former is quite tedious, the sequence of each potential conformer is run in the PDA program to determine whether it is a conformer.

Variant TNF-α proteins may also be identified as being encoded by variant TNF-α nucleic acids. In the case of the nucleic acid, the overall homology of the nucleic acid sequence is commensurate with amino acid homology but takes into account the degeneracy in the genetic code and codon bias of different organisms. Accordingly, the nucleic acid sequence homology may be either lower or higher than that of the protein sequence, with lower homology being preferred.

In a preferred embodiment, a variant TNF-α nucleic acid encodes a variant TNF-α protein. As will be appreciated by those in the art, due to the degeneracy of the genetic code, an extremely large number of nucleic acids may be made, all of which encode the variant TNF-α proteins of the present invention. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids, by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the variant TNF-α.

28

In one embodiment, the nucleic acid homology is determined through hybridization studies. Thus, for example, nucleic acids which hybridize under high stringency to the nucleic acid sequence shown in Figure 6A or its complement and encode a variant TNF-α protein is considered a variant TNF-α gene.

High stringency conditions are known in the art; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, supra, and Tijssen, supra.

The variant TNF-α proteins and nucleic acids of the present invention are recombinant. As used herein, "nucleic acid" may refer to either DNA or RNA, or molecules which contain both deoxy- and ribonucleotides. The nucleic acids include genomic DNA, cDNA and oligonucleotides including sense and anti-sense nucleic acids. Such nucleic acids may also contain modifications in the ribose-phosphate backbone to increase stability and half life of such molecules in physiological environments.

The nucleic acid may be double stranded, single stranded, or contain portions of both double stranded or single stranded sequence. As will be appreciated by those in the art, the depiction of a single strand ("Watson") also defines the sequence of the other strand ("Crick"); thus the sequence depicted in Figure 6 also includes the complement of the sequence. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed in vitro, in general, by the manipulation of nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated variant TNF-α nucleic

acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention.

Similarly, a "recombinant protein" is a protein made using recombinant techniques, i.e. through the expression of a recombinant nucleic acid as depicted above. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. For example, the protein may be isolated or purified away from some or all of the proteins and compounds with which it is normally associated in its wild type host, and thus may be substantially pure. For example, an isolated protein is unaccompanied by at least some of the material with which it is normally associated in its natural state, preferably constituting at least about 0.5%, more preferably at least about 5% by weight of the total protein in a given sample. A substantially pure protein comprises at least about 75% by weight of the total protein, with at least about 80% being preferred, and at least about 90% being particularly preferred. The definition includes the production of a variant TNF-α protein from one organism in a different organism or host cell. Alternatively, the protein may be made at a significantly higher concentration than is normally seen, through the use of a inducible promoter or high expression promoter, such that the protein is made at increased concentration levels. Furthermore, all of the variant TNF-α proteins outlined herein are in a form not normally found in nature, as they contain amino acid substitutions, insertions and deletions, with substitutions being preferred, as discussed below.

Also included within the definition of variant TNF-α proteins of the present invention are amino acid sequence variants of the variant TNF-α sequences outlined herein and shown in the Figures. That is, the variant TNF-α proteins may contain additional variable positions as compared to human TNF-α. These variants fall into one or more of three classes: substitutional, insertional or deletional variants. These variants ordinarily are prepared by site specific mutagenesis of nucleotides in the DNA encoding a variant TNF-α protein, using cassette or PCR mutagenesis or other techniques well known in the art, to produce DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture as outlined above. However, variant TNF-α protein fragments having up to about 100-150 residues may be prepared by *in vitro* synthesis using established techniques. Amino acid sequence variants are characterized by the predetermined nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variation of the variant TNF-α protein amino acid sequence. The variants typically exhibit the same qualitative biological activity as the naturally

occurring analogue, although variants can also be selected which have modified characteristics as will be more fully outlined below.

While the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed variant TNF-α proteins screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants is done using assays of variant TNF-α protein activities.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.

Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the variant TNF-α protein are desired, substitutions are generally made in accordance with the following chart:

Chart I

Original Residue	Exemplary Substitutions
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser, Ala
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those shown in Chart I. For example, substitutions may be made which more significantly affect the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g. lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine.

The variants typically exhibit the same qualitative biological activity and will elicit the same immune response as the original variant TNF-α protein, although variants also are selected to modify the characteristics of the variant TNF-α proteins as needed. Alternatively, the variant may be designed such that the biological activity of the variant TNF-α protein is altered. For example, glycosylation sites may be altered or removed. Similarly, the biological function may be altered; for example, in some instances it may be desirable to have more or less potent TNF-α activity.

In a preferred embodiment, also included within the invention are soluble p55 variant TNF proteins and nucleic acids. In this embodiment, the soluble p55 variant TNF can serve as an antagonist to receptor signaling. By "serving as an antagonist to receptor signaling" herein is meant that the soluble p55 variant TNF proteins preferentially interact with wild-type TNF-α to block or significantly decrease TNF-α receptor activated signaling.

Thus, the computational processing results described above may be used to generate a set of optimized variant p55 protein sequences. Optimized variant p55 protein sequences are generally different from wild-type p55 sequences in at least about 1 variant amino acid.

In a preferred embodiment variant TNF p55 proteins are fused to a human TNF receptor-associated factor (TRAF) trimerization domain. In a preferred embodiment, the C termini of optimized variant TNF p 55 receptors will be fused to TRAF trimerization domains (i.e., leucine zipper motif).

Fusion of trimerization domains from TRAF proteins to TNFR molecules can induce trimerization, resulting in higher avidity for TNFa thereby creating a more potent TNFa inhibitor than the monomeric soluble TNFR. These trimerization domains can be used to induce the trimerization of any protein where this may be desirable, including TNFalpha, TNFbeta, TNF receptor (p55 and p75), and other

members of the TNF receptor family including NGF receptor, CD27, CD30, CD40, fas antigen. Other peptides that are known to form trimeric coiled coils could also be used, including pII (Harbury, Kim and Alber, 1994).

While the description herein is focused on TNF-α variants, as will be appreciated by those in the art, the embodiments and definitions can be applied to soluble p55 variant TNF proteins.

The variant TNF-α proteins and nucleic acids of the invention can be made in a number of ways. Individual nucleic acids and proteins can be made as known in the art and outlined below. Alternatively, libraries of variant TNF-α proteins can be made for testing.

In a preferred embodiment, sets or libraries of variant TNF-α proteins are generated from a probability distribution table. As outlined herein, there are a variety of methods of generating a probability distribution table, including using PDA, sequence alignments, forcefield calculations such as SCMF calculations, etc. In addition, the probability distribution can be used to generate information entropy scores for each position, as a measure of the mutational frequency observed in the library.

In this embodiment, the frequency of each amino acid residue at each variable position in the list is identified. Frequencies can be thresholded, wherein any variant frequency lower than a cutoff is set to zero. This cutoff is preferably 1%, 2%, 5%, 10% or 20%, with 10% being particularly preferred. These frequencies are then built into the variant TNF-α library. That is, as above, these variable positions are collected and all possible combinations are generated, but the amino acid residues that "fill" the library are utilized on a frequency basis. Thus, in a non-frequency based library, a variable position that has 5 possible residues will have 20% of the proteins comprising that variable position with the first possible residue, 20% with the second, etc. However, in a frequency based library, a variable position that has 5 possible residues with frequencies of 10%, 15%, 25%, 30% and 20%, respectively, will have 10% of the proteins comprising that variable position with the first possible residue, 15% of the proteins with the second residue, 25% with the third, etc. As will be appreciated by those in the art, the actual frequency may depend on the method used to actually generate the proteins; for example, exact frequencies may be possible when the proteins are synthesized. However, when the frequency-based primer system outlined below is used, the actual frequencies at each position will vary, as outlined below.

As will be appreciated by those in the art and outlined herein, probability distribution tables can be generated in a variety of ways. In addition to the methods outlined herein, self-consistent mean field (SCMF) methods can be used in the direct generation of probability tables. SCMF is a deterministic computational method that uses a mean field description of rotamer interactions to calculate energies.

A probability table generated in this way can be used to create libraries as described herein. SCMF can be used in three ways: the frequencies of amino acids and rotamers for each amino acid are listed at each position; the probabilities are determined directly from SCMF (see Delarue et la. Pac. Symp. Biocomput. 109-21 (1997), expressly incorporated by reference). In addition, highly variable positions and non-variable positions can be identified. Alternatively, another method is used to determine what sequence is jumped to during a search of sequence space; SCMF is used to obtain an accurate energy for that sequence; this energy is then used to rank it and create a rank-ordered list of sequences (similar to a Monte Carlo sequence list). A probability table showing the frequencies of amino acids at each position can then be calculated from this list (Koehl et al., J. Mol. Biol. 239:249 (1994); Koehl et al., Nat. Struc. Biol. 2:163 (1995); Koehl et al., Curr. Opin. Struct. Biol. 6:222 (1996); Koehl et al., J. Mol. Bio. 293:1183 (1999); Koehl et al., J. Mol. Biol. 293:1161 (1999); Lee J. Mol. Biol. 236:918 (1994); and Vasquez Biopolymers 36:53-70 (1995); all of which are expressly incorporated by reference. Similar methods include, but are not limited to, OPLS-AA (Jorgensen, et al., J. Am. Chem. Soc. (1996), v 118, pp 11225-11236; Jorgensen, W.L.; BOSS, Version 4.1; Yale University: New Haven, CT (1999)); OPLS (Jorgensen, et al., J. Am. Chem. Soc. (1988), v 110, pp 1657ff; Jorgensen, et al., J Am. Chem. Soc. (1990), v 112, pp 4768ff); UNRES (United Residue Forcefield; Liwo, et al., Protein Science (1993), v 2, pp1697-1714; Liwo, et al., Protein Science (1993), v 2, pp1715-1731; Liwo, et al., J. Comp. Chem. (1997), v 18, pp849-873; Liwo, et al., J. Comp. Chem. (1997), v 18, pp874-884; Liwo, et al., J. Comp. Chem. (1998), v 19, pp259-276; Forcefield for Protein Structure Prediction (Liwo, et al., Proc. Natl. Acad. Sci. USA (1999), v 96, pp5482-5485); ECEPP/3 (Liwo et al., J Protein Chem 1994 May;13(4):375-80); AMBER 1.1 force field (Weiner, et al., J. Am. Chem. Soc. v106, pp765-784); AMBER 3.0 force field (U.C. Singh et al., Proc. Natl. Acad. Sci. USA. 82:755-759); CHARMM and CHARMM22 (Brooks, et al., J. Comp. Chem. v4, pp 187-217); cvff3.0 (Dauber-Osguthorpe, et al.,(1988) Proteins: Structure, Function and Genetics, v4,pp31-47); cff91 (Maple, et al., J. Comp. Chem. v15, 162-182); also, the DISCOVER (cvff and cff91) and AMBER forcefields are used in the INSIGHT molecular modeling package (Biosym/MSI, San Diego California) and HARMM is used in the QUANTA molecular modeling package (Biosym/MSI, San Diego California).

In addition, as outlined herein, a preferred method of generating a probability distribution table is through the use of sequence alignment programs. In addition, the probability table can be obtained by a combination of sequence alignments and computational approaches. For example, one can add amino acids found in the alignment of homologous sequences to the result of the computation. Preferable one can add the wild type amino acid identity to the probability table if it is not found in the computation.

As will be appreciated, a variant TNF-α library created by recombining variable positions and/or residues at the variable position may not be in a rank-ordered list. In some embodiments, the entire list may just be made and tested. Alternatively, in a preferred embodiment, the variant TNF-α library is also in the form of a rank ordered list. This may be done for several reasons, including the size of the library is still too big to generate experimentally, or for predictive purposes. This may be done in several ways. In one embodiment, the library is ranked using the scoring functions of PDA to rank the library members. Alternatively, statistical methods could be used. For example, the library may be ranked by frequency score; that is, proteins containing the most of high frequency residues could be ranked higher, etc. This may be done by adding or multiplying the frequency at each variable position to generate a numerical score. Similarly, the library different positions could be weighted and then the proteins scored; for example, those containing certain residues could be arbitrarily ranked.

In a preferred embodiment, the different protein members of the variant TNF-α library may be chemically synthesized. This is particularly useful when the designed proteins are short, preferably less than 150 amino acids in length, with less than 100 amino acids being preferred, and less than 50 amino acids being particularly preferred, although as is known in the art, longer proteins can be made chemically or enzymatically. See for example Wilken et al, Curr. Opin. Biotechnol. 9:412-26 (1998), hereby expressly incorporated by reference.

In a preferred embodiment, particularly for longer proteins or proteins for which large samples are desired, the library sequences are used to create nucleic acids such as DNA which encode the member sequences and which can then be cloned into host cells, expressed and assayed, if desired. Thus, nucleic acids, and particularly DNA, can be made which encodes each member protein sequence. This is done using well known procedures. The choice of codons, suitable expression vectors and suitable host cells will vary depending on a number of factors, and can be easily optimized as needed.

In a preferred embodiment, multiple PCR reactions with pooled oligonucleotides is done, as is generally depicted in the Figures. In this embodiment, overlapping oligonucleotides are synthesized which correspond to the full length gene. Again, these oligonucleotides may represent all of the different amino acids at each variant position or subsets.

In a preferred embodiment, these oligonucleotides are pooled in equal proportions and multiple PCR reactions are performed to create full length sequences containing the combinations of mutations defined by the library. In addition, this may be done using error-prone PCR methods.

35

In a preferred embodiment, the different oligonucleotides are added in relative amounts corresponding to the probability distribution table. The multiple PCR reactions thus result in full length sequences with the desired combinations of mutations in the desired proportions.

The total number of oligonucleotides needed is a function of the number of positions being mutated and the number of mutations being considered at these positions:
(number of oligos for constant positions) + M1 + M2 + M3 +... Mn = (total number of oligos required), where Mn is the number of mutations considered at position n in the sequence.

In a preferred embodiment, each overlapping oligonucleotide comprises only one position to be varied; in alternate embodiments, the variant positions are too close together to allow this and multiple variants per oligonucleotide are used to allow complete recombination of all the possibilities. That is, each oligo can contain the codon for a single position being mutated, or for more than one position being mutated. The multiple positions being mutated must be close in sequence to prevent the oligo length from being impractical. For multiple mutating positions on an oligonucleotide, particular combinations of mutations can be included or excluded in the library by including or excluding the oligonucleotide encoding that combination. For example, as discussed herein, there may be correlations between variable regions; that is, when position X is a certain residue, position Y must (or must not) be a particular residue. These sets of variable positions are sometimes referred to herein as a "cluster". When the clusters are comprised of residues close together, and thus can reside on one oligonucleotide primer, the clusters can be set to the "good" correlations, and eliminate the bad combinations that may decrease the effectiveness of the library. However, if the residues of the cluster are far apart in sequence, and thus will reside on different oligonucleotides for synthesis, it may be desirable to either set the residues to the "good" correlation, or eliminate them as variable residues entirely. In an alternative embodiment, the library may be generated in several steps, so that the cluster mutations only appear together. This procedure, i.e. the procedure of identifying mutation clusters and either placing them on the same oligonucleotides or eliminating them from the library or library generation in several steps preserving clusters, can considerably enrich the experimental library with properly folded protein. Identification of clusters can be carried out by a number of ways, e.g. by using known pattern recognition methods, comparisons of frequencies of occurrence of mutations or by using energy analysis of the sequences to be experimentally generated (for example, if the energy of interaction is high, the positions are correlated). These correlations may be positional correlations (e.g. variable positions 1 and 2 always change together or never change together) or sequence correlations (e.g. if there is residue A at position 1, there is always residue B at position 2). See: Pattern discovery in Biomolecular Data: Tools, Techniques, and Applications; edited by Jason T.L. Wang, Bruce A. Shapiro, Dennis Shasha. New York: Oxford University, 1999; Andrews, Harry C. Introduction to mathematical techniques in pattern recognition; New York, Wiley-Interscience [1972];

Applications of Pattern Recognition; Editor, K.S. Fu. Boca Raton, Fla. CRC Press, 1982; Genetic Algorithms for Pattern Recognition; edited by Sankar K. Pal, Paul P. Wang. Boca Raton: CRC Press, c1996; Pandya, Abhijit S., Pattern recognition with neural networks in C++ / Abhijit S. Pandya, Robert B. Macy. Boca Raton, Fla.: CRC Press, 1996; Handbook of pattern recognition & computer vision / edited by C.H. Chen, L.F. Pau, P.S.P. Wang. 2nd ed. Singapore; River Edge, N.J.: World Scientific, c1999; Friedman, Introduction to Pattern Recognition: Statistical, Structural, Neural, and Fuzy Logic Approaches; River Edge, N.J.: World Scientific, c1999, Series title: Series in machine perception and artificial intelligence; vol. 32; all of which are expressly incorporated by reference. In addition, programs used to search for consensus motifs can be used as well.

In addition, correlations and shuffling can be fixed or optimized by altering the design of the oligonucleotides; that is, by deciding where the oligonucleotides (primers) start and stop (e.g. where the sequences are "cut"). The start and stop sites of oligos can be set to maximize the number of clusters that appear in single oligonucleotides, thereby enriching the library with higher scoring sequences. Different oligonucleotide start and stop site options can be computationally modeled and ranked according to number of clusters that are represented on single oligos, or the percentage of the resulting sequences consistent with the predicted library of sequences.

The total number of oligonucleotides required increases when multiple mutable positions are encoded by a single oligonucleotide. The annealed regions are the ones that remain constant, i.e. have the sequence of the reference sequence.

Oligonucleotides with insertions or deletions of codons can be used to create a library expressing different length proteins. In particular computational sequence screening for insertions or deletions can result in secondary libraries defining different length proteins, which can be expressed by a library of pooled oligonucleotide of different lengths.

In a preferred embodiment, the variant TNF-α library is done by shuffling the family (e.g. a set of variants); that is, some set of the top sequences (if a rank-ordered list is used) can be shuffled, either with or without error-prone PCR. "Shuffling" in this context means a recombination of related sequences, generally in a random way. It can include "shuffling" as defined and exemplified in U.S. Patent Nos. 5,830,721; 5,811,238; 5,605,793; 5,837,458 and PCT US/19256, all of which are expressly incorporated by reference in their entirety. This set of sequences can also be an artificial set; for example, from a probability table (for example generated using SCMF) or a Monte Carlo set. Similarly, the "family" can be the top 10 and the bottom 10 sequences, the top 100 sequence, etc. This may also be done using error-prone PCR.

37

Thus, in a preferred embodiment, in silico shuffling is done using the computational methods described herein. That is, starting with either two libraries or two sequences, random recombinations of the sequences can be generated and evaluated.

In a preferred embodiment, error-prone PCR is done to generate the variant TNF-α library. See U.S. Patent Nos. 5,605,793, 5,811,238, and 5,830,721, all of which are hereby incorporated by reference. This can be done on the optimal sequence or on top members of the library, or some other artificial set or family. In this embodiment, the gene for the optimal sequence found in the computational screen of the primary library can be synthesized. Error prone PCR is then performed on the optimal sequence gene in the presence of oligonucleotides that code for the mutations at the variant positions of the library (bias oligonucleotides). The addition of the oligonucleotides will create a bias favoring the incorporation of the mutations in the library. Alternatively, only oligonucleotides for certain mutations may be used to bias the library.

In a preferred embodiment, gene shuffling with error prone PCR can be performed on the gene for the optimal sequence, in the presence of bias oligonucleotides, to create a DNA sequence library that reflects the proportion of the mutations found in the variant TNF-α library. The choice of the bias oligonucleotides can be done in a variety of ways; they can chosen on the basis of their frequency, i.e. oligonucleotides encoding high mutational frequency positions can be used; alternatively, oligonucleotides containing the most variable positions can be used, such that the diversity is increased; if the secondary library is ranked, some number of top scoring positions can be used to generate bias oligonucleotides; random positions may be chosen; a few top scoring and a few low scoring ones may be chosen; etc. What is important is to generate new sequences based on preferred variable positions and sequences.

In a preferred embodiment, PCR using a wild type gene or other gene can be used, as is schematically depicted in the Figures. In this embodiment, a starting gene is used; generally, although this is not required, the gene is usually the wild type gene. In some cases it may be the gene encoding the global optimized sequence, or any other sequence of the list, or a consensus sequence obtained e.g. from aligning homologous sequences from different organisms. In this embodiment, oligonucleotides are used that correspond to the variant positions and contain the different amino acids of the library. PCR is done using PCR primers at the termini, as is known in the art. This provides two benefits; the first is that this generally requires fewer oligonucleotides and can result in fewer errors. In addition, it has experimental advantages in that if the wild type gene is used, it need not be synthesized.

In addition, there are several other techniques that can be used, as exemplified in the figures. In a preferred embodiment, ligation of PCR products is done.

In a preferred embodiment, a variety of additional steps may be done to the variant TNF-α library; for example, further computational processing can occur, different variant TNF-α libraries can be recombined, or cutoffs from different libraries can be combined. In a preferred embodiment, a variant TNF-α library may be computationally remanipulated to form an additional variant TNF-α library (sometimes referred to herein as "tertiary libraries"). For example, any of the variant TNF-α library sequences may be chosen for a second round of PDA, by freezing or fixing some or all of the changed positions in the first library. Alternatively, only changes seen in the last probability distribution table are allowed. Alternatively, the stringency of the probability table may be altered, either by increasing or decreasing the cutoff for inclusion. Similarly, the variant TNF-α library may be recombined experimentally after the first round; for example, the best gene/genes from the first screen may be taken and gene assembly redone (using techniques outlined below, multiple PCR, error prone PCR, shuffling, etc.). Alternatively, the fragments from one or more good gene(s) to change probabilities at some positions. This biases the search to an area of sequence space found in the first round of computational and experimental screening.

In a preferred embodiment, a tertiary library can be generated from combining different variant TNF-α libraries. For example, a probability distribution table from a first variant TNF-α library can be generated and recombined, either computationally or experimentally, as outlined herein. A PDA variant TNF-α library may be combined with a sequence alignment variant TNF-α library, and either recombined (again, computationally or experimentally) or just the cutoffs from each joined to make a new tertiary library. The top sequences from several libraries can be recombined. Sequences from the top of a library can be combined with sequences from the bottom of the library to more broadly sample sequence space, or only sequences distant from the top of the library can be combined. Variant TNF-α libraries that analyzed different parts of a protein can be combined to a tertiary library that treats the combined parts of the protein.

In a preferred embodiment, a tertiary library can be generated using correlations in a variant TNF-α library. That is, a residue at a first variable position may be correlated to a residue at second variable position (or correlated to residues at additional positions as well). For example, two variable positions may sterically or electrostatically interact, such that if the first residue is X, the second residue must be Y. This may be either a positive or negative correlation.

Using the nucleic acids of the present invention which encode a variant TNF-α protein, a variety of expression vectors are made. The expression vectors may be either self-replicating

extrachromosomal vectors or vectors which integrate into a host genome. Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the variant TNF-α protein. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation.

In a preferred embodiment, when the endogenous secretory sequence leads to a low level of secretion of the naturally occurring protein or of the variant TNF-α protein, a replacement of the naturally occurring secretory leader sequence is desired. In this embodiment, an unrelated secretory leader sequence is operably linked to a variant TNF-α encoding nucleic acid leading to increased protein secretion. Thus, any secretory leader sequence resulting in enhanced secretion of the variant TNF-α protein, when compared to the secretion of TNF-α and its secretory sequence, is desired. Suitable secretory leader sequences that lead to the secretion of a protein are know in the art.

In another preferred embodiment, a secretory leader sequence of a naturally occurring protein or a protein is removed by techniques known in the art and subsequent expression results in intracellular accumulation of the recombinant protein.

Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. The transcriptional and translational regulatory nucleic acid will generally be appropriate to the host cell used to express the fusion protein; for example, transcriptional and translational regulatory nucleic acid sequences from *Bacillus* are preferably used to express the fusion protein in *Bacillus*. Numerous types of appropriate expression vectors, and suitable regulatory sequences are known in the art for a variety of host cells.

40

In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention. In a preferred embodiment, the promoters are strong promoters, allowing high expression in cells, particularly mammalian cells, such as the CMV promoter, particularly in combination with a Tet regulatory element.

In addition, the expression vector may comprise additional elements. For example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct. The integrating vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art.

In addition, in a preferred embodiment, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used.

A preferred expression vector system is a retroviral vector system such as is generally described in PCT/US97/01019 and PCT/US97/01048; both of which are hereby expressly incorporated by reference.

In a preferred embodiment, the expression vector comprises the components described above and a gene encoding a variant TNF-α protein. As will be appreciated by those in the art, all combinations are possible and accordingly, as used herein, the combination of components, comprised by one or more vectors, which may be retroviral or not, is referred to herein as a "vector composition".

The variant TNF-α nucleic acids are introduced into the cells either alone or in combination with an expression vector. By "introduced into " or grammatical equivalents herein is meant that the nucleic

acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type, discussed below. Exemplary methods include $CaPO_4$ precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The variant TNF-α nucleic acids may stably integrate into the genome of the host cell (for example, with retroviral introduction, outlined below), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.).

The variant TNF-α proteins of the present invention are produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a variant TNF-α protein, under the appropriate conditions to induce or cause expression of the variant TNF-α protein. The conditions appropriate for variant TNF-α protein expression will vary with the choice of the expression vector and the host cell, and will be easily ascertained by one skilled in the art through routine experimentation. For example, the use of constitutive promoters in the expression vector will require optimizing the growth and proliferation of the host cell, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important. For example, the baculoviral systems used in insect cell expression are lytic viruses, and thus harvest time selection can be crucial for product yield.

Appropriate host cells include yeast, bacteria, archebacteria, fungi, and insect and animal cells, including mammalian cells. Of particular interest are *Drosophila melangaster* cells, *Saccharomyces cerevisiae* and other yeasts, *E. coli, Bacillus subtilis*, SF9 cells, C129 cells, 293 cells, Neurospora, BHK, CHO, COS, *Pichia Pastoris*, etc.

In a preferred embodiment, the variant TNF-α proteins are expressed in mammalian cells. Mammalian expression systems are also known in the art, and include retroviral systems. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence for the fusion protein into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, using a located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element (enhancer element), typically located within 100 to 200 base pairs upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as mammalian promoters are the promoters from mammalian viral genes, since the viral genes are often highly expressed and have a broad host range. Examples

include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter, herpes simplex virus promoter, and the CMV promoter.

Typically, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of transcription terminator and polyadenlytion signals include those derived form SV40.

The methods of introducing exogenous nucleic acid into mammalian hosts, as well as other hosts, is well known in the art, and will vary with the host cell used. Techniques include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, viral infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei. As outlined herein, a particularly preferred method utilizes retroviral infection, as outlined in PCT US97/01019, incorporated by reference.

As will be appreciated by those in the art, the type of mammalian cells used in the present invention can vary widely. Basically, any mammalian cells may be used, with mouse, rat, primate and human cells being particularly preferred, although as will be appreciated by those in the art, modifications of the system by pseudotyping allows all 'eukaryotic cells to be used, preferably higher eukaryotes. As is more fully described below, a screen will be set up such that the cells exhibit a selectable phenotype in the presence of a bioactive peptide. As is more fully described below, cell types implicated in a wide variety of disease conditions are particularly useful, so long as a suitable screen may be designed to allow the selection of cells that exhibit an altered phenotype as a consequence of the presence of a peptide within the cell.

Accordingly, suitable cell types include, but are not limited to, tumor cells of all types (particularly melanoma, myeloid leukemia, carcinomas of the lung, breast, ovaries, colon, kidney, prostate, pancreas and testes), cardiomyocytes, endothelial cells, epithelial cells, lymphocytes (T-cell and B cell) , mast cells, eosinophils, vascular intimal cells, hepatocytes, leukocytes including mononuclear leukocytes, stem cells such as haemopoetic, neural, skin, lung, kidney, liver and myocyte stem cells (for use in screening for differentiation and de-differentiation factors), osteoclasts, chondrocytes and other connective tissue cells, keratinocytes, melanocytes, liver cells, kidney cells, and adipocytes. Suitable cells also include known research cells, including, but not limited to, Jurkat T cells, NIH3T3 cells, CHO, Cos, etc. See the ATCC cell line catalog, hereby expressly incorporated by reference.

In one embodiment, the cells may be additionally genetically engineered, that is, contain exogeneous nucleic acid other than the variant TNF-α nucleic acid.

In a preferred embodiment, the variant TNF-α proteins are expressed in bacterial systems. Bacterial expression systems are well known in the art.

A suitable bacterial promoter is any nucleic acid sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of the coding sequence of the variant TNF-α protein into mRNA. A bacterial promoter has a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose and maltose, and sequences derived from biosynthetic enzymes such as tryptophan. Promoters from bacteriophage may also be used and are known in the art. In addition, synthetic promoters and hybrid promoters are also useful; for example, the *tac* promoter is a hybrid of the *trp* and *lac* promoter sequences. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription.

In addition to a functioning promoter sequence, an efficient ribosome binding site is desirable. In *E. coli*, the ribosome binding site is called the Shine-Delgarno (SD) sequence and includes an initiation codon and a sequence 3-9 nucleotides in length located 3 - 11 nucleotides upstream of the initiation codon.

The expression vector may also include a signal peptide sequence that provides for secretion of the variant TNF-α protein in bacteria. The signal sequence typically encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell, as is well known in the art. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). For expression in bacteria, usually bacterial secretory leader sequences, operably linked to a variant TNF-α encoding nucleic acid, are preferred.

The bacterial expression vector may also include a selectable marker gene to allow for the selection of bacterial strains that have been transformed. Suitable selection genes include genes which render the bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin and tetracycline. Selectable markers also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

These components are assembled into expression vectors. Expression vectors for bacteria are well known in the art, and include vectors for *Bacillus subtilis, E. coli, Streptococcus cremoris*, and *Streptococcus lividans*, among others.

The bacterial expression vectors are transformed into bacterial host cells using techniques well known in the art, such as calcium chloride treatment, electroporation, and others.

In one embodiment, variant TNF-α proteins are produced in insect cells. Expression vectors for the transformation of insect cells, and in particular, baculovirus-based expression vectors, are well known in the art.

In a preferred embodiment, variant TNF-α protein is produced in yeast cells. Yeast expression systems are well known in the art, and include expression vectors for *Saccharomyces cerevisiae, Candida albicans* and *C. maltosa, Hansenula polymorpha, Kluyveromyces fragilis* and *K. lactis, Pichia guillerimondii* and *P. pastoris, Schizosaccharomyces pombe*, and *Yarrowia lipolytica*. Preferred promoter sequences for expression in yeast include the inducible GAL1,10 promoter, the promoters from alcohol dehydrogenase, enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase, hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, pyruvate kinase, and the acid phosphatase gene. Yeast selectable markers include ADE2, HIS4, LEU2, TRP1, and ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions.

In addition, the variant TNF-α polypeptides of the invention may be further fused to other proteins, if desired, for example to increase expression or stabilize the protein.

In one embodiment, the variant TNF-α nucleic acids, proteins and antibodies of the invention are labeled with a label other than the scaffold. By "labeled" herein is meant that a compound has at least one element, isotope or chemical compound attached to enable the detection of the compound. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) immune labels, which may be antibodies or antigens; and c) colored or fluorescent dyes. The labels may be incorporated into the compound at any position.

Once made, the variant TNF-α proteins may be covalently modified. Covalent and non-covalent modifications of the protein are thus included within the scope of the present invention. Such modifications may be introduced into a variant TNF-α polypeptide by reacting targeted amino acid residues of the polypeptide with an organic derivatizing agent that is capable of reacting with selected

side chains or terminal residues.

One type of covalent modification includes reacting targeted amino acid residues of a variant TNF-α polypeptide with an organic derivatizing agent that is capable of reacting with selected side chains or the N-or C-terminal residues of a variant TNF-α polypeptide. Derivatization with bifunctional agents is useful, for instance, for crosslinking a variant TNF-α protein to a water-insoluble support matrix or surface for use in the method for purifying anti-variant TNF-α antibodies or screening assays, as is more fully described below. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidyl-propionate), bifunctional maleimides such as bis-N-maleimido-1,8-octane and agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate.

Other modifications include deamidation of glutaminyl and asparaginyl residues to the corresponding glutamyl and aspartyl residues, respectively, hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the "-amino groups of lysine, arginine, and histidine side chains [T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)], acetylation of the N-terminal amine, and amidation of any C-terminal carboxyl group.

Another type of covalent modification of the variant TNF-α polypeptide included within the scope of this invention comprises altering the native glycosylation pattern of the polypeptide. "Altering the native glycosylation pattern" is intended for purposes herein to mean deleting one or more carbohydrate moieties found in native sequence variant TNF-α polypeptide, and/or adding one or more glycosylation sites that are not present in the native sequence variant TNF-α polypeptide.

Addition of glycosylation sites to variant TNF-α polypeptides may be accomplished by altering the amino acid sequence thereof. The alteration may be made, for example, by the addition of, or substitution by, one or more serine or threonine residues to the native sequence variant TNF-α polypeptide (for O-linked glycosylation sites). The variant TNF-α amino acid sequence may optionally be altered through changes at the DNA level, particularly by mutating the DNA encoding the variant TNF-α polypeptide at preselected bases such that codons are generated that will translate into the desired amino acids.

Another means of increasing the number of carbohydrate moieties on the variant TNF-α polypeptide is by chemical or enzymatic coupling of glycosides to the polypeptide. Such methods are described in

46

the art, e.g., in WO 87/05330 published 11 September 1987, and in Aplin and Wriston, CRC Crit. Rev. Biochem., pp. 259-306 (1981).

Removal of carbohydrate moieties present on the variant TNF-α polypeptide may be accomplished chemically or enzymatically or by mutational substitution of codons encoding for amino acid residues that serve as targets for glycosylation. Chemical deglycosylation techniques are known in the art and described, for instance, by Hakimuddin, et al., Arch. Biochem. Biophys., 259:52 (1987) and by Edge et al., Anal. Biochem., 118:131 (1981). Enzymatic cleavage of carbohydrate moieties on polypeptides can be achieved by the use of a variety of endo-and exo-glycosidases as described by Thotakura et al., Meth. Enzymol., 138:350 (1987).

Such derivatized moieties may improve the solubility, absorption, permeability across the blood brain barrier biological half life, and the like. Such moieties or modifications of variant TNF-α polypeptides may alternatively eliminate or attenuate any possible undesirable side effect of the protein and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 16th ed., Mack Publishing Co., Easton, Pa. (1980).

Another type of covalent modification of variant TNF-α comprises linking the variant TNF-α polypeptide to one of a variety of nonproteinaceous polymers, e.g., polyethylene glycol, polypropylene glycol, or polyoxyalkylenes, in the manner set forth in U.S. Patent Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192 or 4,179,337.

Variant TNF-α polypeptides of the present invention may also be modified in a way to form chimeric molecules comprising a variant TNF-α polypeptide fused to another, heterologous polypeptide or amino acid sequence. In one embodiment, such a chimeric molecule comprises a fusion of a variant TNF-α polypeptide with a tag polypeptide which provides an epitope to which an anti-tag antibody can selectively bind. The epitope tag is generally placed at the amino-or carboxyl-terminus of the variant TNF-α polypeptide. The presence of such epitope-tagged forms of a variant TNF-α polypeptide can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the variant TNF-α polypeptide to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. In an alternative embodiment, the chimeric molecule may comprise a fusion of a variant TNF-α polypeptide with an immunoglobulin or a particular region of an immunoglobulin. For a bivalent form of the chimeric molecule, such a fusion could be to the Fc region of an IgG molecule.

Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its

antibody 12CA5 [Field et al., Mol. Cell. Biol. 8:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)]; and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky et al., Protein Engineering, 3(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp et al., BioTechnology 6:1204-1210 (1988)]; the KT3 epitope peptide [Martin et al., Science 255:192-194 (1992)]; tubulin epitope peptide [Skinner et al., J. Biol. Chem. 266:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. U.S.A. 87:6393-6397 (1990)].

In a preferred embodiment, the variant TNF-α protein is purified or isolated after expression. Variant TNF-α proteins may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. For example, the variant TNF-α protein may be purified using a standard anti-library antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. For general guidance in suitable purification techniques, see Scopes, R., Protein Purification, Springer-Verlag, NY (1982). The degree of purification necessary will vary depending on the use of the variant TNF-α protein. In some instances no purification will be necessary.

Once made, the variant TNF-α proteins and nucleic acids of the invention find use in a number of applications. In a preferred embodiment, the variant TNF-α proteins are administered to a patient to treat an TNF-α related disorder.

By "TNF-α related disorder" or "TNF-α responsive disorder" or "condition" herein is meant a disorder that can be ameliorated by the administration of a pharmaceutical composition comprising a variant TNF-α protein, including, but not limited to, inflammatory and immunological disorders. In a preferred embodiment, the variant TNF-α protein is used to treat rheumatoid arthritis.

In a preferred embodiment, a therapeutically effective dose of a variant TNF-α protein is administered to a patient in need of treatment. By "therapeutically effective dose" herein is meant a dose that produces the effects for which it is administered. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques. In a preferred embodiment, dosages of about 5 μg/kg are used, administered either intraveneously or subcutaneously. As is known in the art, adjustments for variant TNF-α protein degradation, systemic versus localized delivery, and rate of new protease synthesis, as well as the age, body weight, general

health, sex, diet, time of administration, drug interaction and the severity of the condition may be necessary, and will be ascertainable with routine experimentation by those skilled in the art.

A "patient" for the purposes of the present invention includes both humans and other animals, particularly mammals, and organisms. Thus the methods are applicable to both human therapy and veterinary applications. In the preferred embodiment the patient is a mammal, and in the most preferred embodiment the patient is human.

The term "treatment" in the instant invention is meant to include therapeutic treatment, as well as prophylactic, or suppressive measures for the disease or disorder. Thus, for example, successful administration of a variant TNF-α protein prior to onset of the disease results in "treatment" of the disease. As another example, successful administration of a variant TNF-α protein after clinical manifestation of the disease to combat the symptoms of the disease comprises "treatment" of the disease. "Treatment" also encompasses administration of a variant TNF-α protein after the appearance of the disease in order to eradicate the disease. Successful administration of an agent after onset and after clinical symptoms have developed, with possible abatement of clinical symptoms and perhaps amelioration of the disease, comprises "treatment" of the disease.

Those "in need of treatment" include mammals already having the disease or disorder, as well as those prone to having the disease or disorder, including those in which the disease or disorder is to be prevented.

In another embodiment, a therapeutically effective dose of a variant TNF-α protein, a variant TNF-α gene, or a variant TNF-α antibody is administered to a patient having a disease involving inappropriate expression of TNF-α. A "disease involving inappropriate expression of at TNF-α" within the scope of the present invention is meant to include diseases or disorders characterized by aberrant TNF-α, either by alterations in the amount of TNF-α present or due to the presence of mutant TNF-α. An overabundance may be due to any cause, including, but not limited to, overexpression at the molecular level, prolonged or accumulated appearance at the site of action, or increased activity of TNF-α relative to normal. Included within this definition are diseases or disorders characterized by a reduction of TNF-α. This reduction may be due to any cause, including, but not limited to, reduced expression at the molecular level, shortened or reduced appearance at the site of action, mutant forms of TNF-α, or decreased activity of TNF-α relative to normal. Such an overabundance or reduction of TNF-α can be measured relative to normal expression, appearance, or activity of TNF-α according to, but not limited to, the assays described and referenced herein.

49

The administration of the variant TNF-α proteins of the present invention, preferably in the form of a sterile aqueous solution, can be done in a variety of ways, including, but not limited to, orally, subcutaneously, intravenously, intranasally, transdermally, intraperitoneally, intramuscularly, intrapulmonary, vaginally, rectally, or intraocularly. In some instances, for example, in the treatment of wounds, inflammation, etc., the variant TNF-α protein may be directly applied as a solution or spray. Depending upon the manner of introduction, the pharmaceutical composition may be formulated in a variety of ways. The concentration of the therapeutically active variant TNF-α protein in the formulation may vary from about 0.1 to 100 weight %. In another preferred embodiment, the concentration of the variant TNF-α protein is in the range of 0.003 to 1.0 molar, with dosages from 0.03, 0.05, 0.1, 0.2, and 0.3 millimoles per kilogram of body weight being preferred.

The pharmaceutical compositions of the present invention comprise a variant TNF-α protein in a form suitable for administration to a patient. In the preferred embodiment, the pharmaceutical compositions are in a water soluble form, such as being present as pharmaceutically acceptable salts, which is meant to include both acid and base addition salts. "Pharmaceutically acceptable acid addition salt" refers to those salts that retain the biological effectiveness of the free bases and that are not biologically or otherwise undesirable, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like, and organic acids such as acetic acid, propionic acid, glycolic acid, pyruvic acid, oxalic acid, maleic acid, malonic acid, succinic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid and the like. "Pharmaceutically acceptable base addition salts" include those derived from inorganic bases such as sodium, potassium, lithium, ammonium, calcium, magnesium, iron, zinc, copper, manganese, aluminum salts and the like. Particularly preferred are the ammonium, potassium, sodium, calcium, and magnesium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines and basic ion exchange resins, such as isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, and ethanolamine.

The pharmaceutical compositions may also include one or more of the following: carrier proteins such as serum albumin; buffers such as NaOAc; fillers such as microcrystalline cellulose, lactose, corn and other starches; binding agents; sweeteners and other flavoring agents; coloring agents; and polyethylene glycol. Additives are well known in the art, and are used in a variety of formulations.

In a further embodiment, the variant TNF-α proteins are added in a micellular formulation; see U.S. Patent No.5,833,948, hereby expressly incorporated by reference in its entirety.

Combinations of pharmaceutical compositions may be administered. Moreover, the compositions may be administered in combination with other therapeutics.

In one embodiment provided herein, antibodies, including but not limited to monoclonal and polyclonal antibodies, are raised against variant TNF-α proteins using methods known in the art. In a preferred embodiment, these anti-variant TNF-α antibodies are used for immunotherapy. Thus, methods of immunotherapy are provided. By "immunotherapy" is meant treatment of an TNFα related disorders with an antibody raised against a variant TNF-α protein. As used herein, immunotherapy can be passive or active. Passive immunotherapy, as defined herein, is the passive transfer of antibody to a recipient (patient). Active immunization is the induction of antibody and/or T-cell responses in a recipient (patient). Induction of an immune response can be the consequence of providing the recipient with a variant TNF-α protein antigen to which antibodies are raised. As appreciated by one of ordinary skill in the art, the variant TNF-α protein antigen may be provided by injecting a variant TNF-α polypeptide against which antibodies are desired to be raised into a recipient, or contacting the recipient with a variant TNF-α protein encoding nucleic acid, capable of expressing the variant TNF-α protein antigen, under conditions for expression of the variant TNF-α protein antigen.

In another preferred embodiment, a therapeutic compound is conjugated to an antibody, preferably an anti-variant TNF-α protein antibody. The therapeutic compound may be a cytotoxic agent. In this method, targeting the cytotoxic agent to tumor tissue or cells, results in a reduction in the number of afflicted cells, thereby reducing symptoms associated with cancer, and variant TNF-α protein related disorders. Cytotoxic agents are numerous and varied and include, but are not limited to, cytotoxic drugs or toxins or active fragments of such toxins. Suitable toxins and their corresponding fragments include diptheria A chain, exotoxin A chain, ricin A chain, abrin A chain, curcin, crotin, phenomycin, enomycin and the like. Cytotoxic agents also include radiochemicals made by conjugating radioisotopes to antibodies raised against cell cycle proteins, or binding of a radionuclide to a chelating agent that has been covalently attached to the antibody.

In a preferred embodiment, variant TNF-α proteins are administered as therapeutic agents, and can be formulated as outlined above. Similarly, variant TNF-α genes (including both the full-length sequence, partial sequences, or regulatory sequences of the variant TNF-α coding regions) can be administered in gene therapy applications, as is known in the art. These variant TNF-α genes can include antisense applications, either as gene therapy (i.e. for incorporation into the genome) or as antisense compositions, as will be appreciated by those in the art.

In a preferred embodiment, the nucleic acid encoding the variant TNF-α proteins may also be used in gene therapy. In gene therapy applications, genes are introduced into cells in order to achieve *in vivo*

synthesis of a therapeutically effective genetic product, for example for replacement of a defective gene. "Gene therapy" includes both conventional gene therapy where a lasting effect is achieved by a single treatment, and the administration of gene therapeutic agents, which involves the one time or repeated administration of a therapeutically effective DNA or mRNA. Antisense RNAs and DNAs can be used as therapeutic agents for blocking the expression of certain genes *in vivo*. It has already been shown that short antisense oligonucleotides can be imported into cells where they act as inhibitors, despite their low intracellular concentrations caused by their restricted uptake by the cell membrane. [Zamecnik et al., Proc. Natl. Acad. Sci. U.S.A. 83:4143-4146 (1986)]. The oligonucleotides can be modified to enhance their uptake, e.g. by substituting their negatively charged phosphodiester groups by uncharged groups.

There are a variety of techniques available for introducing nucleic acids into viable cells. The techniques vary depending upon whether the nucleic acid is transferred into cultured cells *in vitro*, or *in vivo* in the cells of the intended host. Techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, the calcium phosphate precipitation method, etc. The currently preferred *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors and viral coat protein-liposome mediated transfection [Dzau et al., Trends in Biotechnology 11:205-210 (1993)]. In some situations it is desirable to provide the nucleic acid source with an agent that targets the target cells, such as an antibody specific for a cell surface membrane protein or the target cell, a ligand for a receptor on the target cell, etc. Where liposomes are employed, proteins which bind to a cell surface membrane protein associated with endocytosis may be used for targeting and/or to facilitate uptake, e.g. capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half-life. The technique of receptor-mediated endocytosis is described, for example, by Wu et al., J. Biol. Chem. 262:4429-4432 (1987); and Wagner et al., Proc. Natl. Acad. Sci. U.S.A. 87:3410-3414 (1990). For review of gene marking and gene therapy protocols see Anderson et al., Science 256:808-813 (1992).

In a preferred embodiment, variant TNF-α genes are administered as DNA vaccines, either single genes or combinations of variant TNF-α genes. Naked DNA vaccines are generally known in the art. Brower, Nature Biotechnology, 16:1304-1305 (1998). Methods for the use of genes as DNA vaccines are well known to one of ordinary skill in the art, and include placing a variant TNF-α gene or portion of a variant TNF-α gene under the control of a promoter for expression in a patient in need of treatment. The variant TNF-α gene used for DNA vaccines can encode full-length variant TNF-α proteins, but more preferably encodes portions of the variant TNF-α proteins including peptides derived from the variant TNF-α protein. In a preferred embodiment a patient is immunized with a DNA vaccine

comprising a plurality of nucleotide sequences derived from a variant TNF-α gene. Similarly, it is possible to immunize a patient with a plurality of variant TNF-α genes or portions thereof as defined herein. Without being bound by theory, expression of the polypeptide encoded by the DNA vaccine, cytotoxic T-cells, helper T-cells and antibodies are induced which recognize and destroy or eliminate cells expressing TNF-α proteins.

In a preferred embodiment, the DNA vaccines include a gene encoding an adjuvant molecule with the DNA vaccine. Such adjuvant molecules include cytokines that increase the immunogenic response to the variant TNF-α polypeptide encoded by the DNA vaccine. Additional or alternative adjuvants are known to those of ordinary skill in the art and find use in the invention.

All references cited herein are incorporated by reference in their entirety.

EXAMPLES

Example 1

Protocol for TNFα Library Expression and Purification

Methods:

1) Overnight culture preparation:

Competent Tuner(DE3)pLysS cells in 96 well-PCR plates were transformed with 1ul of TNFa library DNAs and spread on LB agar plates with 34 μg/ml chloramphenicol and 100 μg/ml ampicillin. After an overnight growth at 37°C, a colony was picked from each plate in 1.5 ml of CG media with 34 μg/ml chloramphenicol and 100 μg/ml ampicilline kept in 96 deep well block. The block was shaken at 250 rpm at 37°C overnight.

2) Expression:

Colonies were picked from the plate into 5 ml CG media (34 μg/ml chloramphenicol and 100 μg/ml ampicillin) in 24-well block and grown at 37°C at 250 rpm until OD600 0.6 were reached, at which time IPTG was added to each well to 1μM concentration. The culture was grown 4 extra hours

3) Lysis:

The 24-well block was centrifuged at 3000 rpm for 10 minutes. The pellets were resuspended in 700 ul of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole). After freezing at −80°C for 20 minutes and thawing at 37°C twice, $MgCl_2$ was added to 10 mM, and DNase I to 75 μg/ml. The mixure was incubated at 37°C for 30 minutes.

4) Ni NTA column purification:

Purification was carried out following Qiagen Ni NTA spin column purification protocol for native condition. The purified protein was dialyzed against 1 X PBS for 1 hour at 4°C four times. Dialyzed protein was filter sterilized, using Millipore multiscreenGV filter plate to allow the addition of protein to the sterile mammalian cell culture assay later on.

5) Quantification:

Purified protein was quantified by SDS PAGE, followed by Coomassie stain, and by Kodak digital image densitometry.

6) TNF-α Activity Assay assay:

The activity of variant TNF-α protein samples was tested using Vybrant Assay Kit and Caspase Assay kit. Sytox Green nucleic acid stain is used to detect TNF-induced cell permeability in Actinomycin-D sensitized cell line. Upon binding to cellular nucleic acids, the stain exhibits a large fluorescence enhancement, which is then measured . This stain is excluded from live cells but penetrates cells with compromised membranes.

Caspase assay is a fluorimetric assay, which can differentiate between apoptosis and necrosis in the cells. This kit measures the caspase activity, triggered during apoptosis of the cells.

A) Materials:

Cell Line:	WEHI Var-13 Cell line from ATCC
Media:	RPMI Complete media with 10% FBS.
Vybrant TNF Kit:	Cat # V-23100 ; Molecular Probes
	Kit contains SYTOX Green nucleic acid stain (500 mM solution)
	and Actinomycin D (1 mg/mL)
Caspase Assay Kit:	Cat # 3 005 372; Roche ·
	Kit contains substrate stock solution (500 uM)
	and incubation buffer
TNF-α Standard stock:	10 ug/mL stock of h-TNF-α from R & D
Unknown Samples:	In house TNF-α library samples
96-well Plates :	1mL deep well and 250 uL wells
Micro plate Reader	

B) Method:

54

Plate WEHI cells at 2.5 x 10⁵ cells/mL, 24 hrs prior to the assay; (100 μL/well for the Sytox assay and 50 μL/well for the Caspase assay).

On the day of the experiment, prepare assay media as follows:

1) Assay Media for Sytox Assay (1X): Prepare assay medium by diluting the concentrated Sytox Green stain and the concentrated actinomycin D solution 500-fold into RPMI, to a final concentration of 10 μM Sytox and 2 μg/mL actinomycin D.

  10 mL complete RPMI medium

  20 μL SYTOX Green

  20 μL actinomycin D

2) Prepare Assay Media for Caspase Assay (1X):

  10 mL complete RPMI medium .

  20 uL Actinomycin D (2 μg/mL final conc.)

3) Prepare Assay Media for samples: Sytox Assay (2X):

  14 mL complete RPMI medium

  56 μL SYTOX Green Nuclei acid stain

  56 μL actinomycin D

4) Prepare Assay Media: (2X): For samples: Caspase assay

  14 mL complete RPMI medium

  56 μL actinomycin D

5) Set up and Run a Standard Curve Dilution:

  TNF-α Std. stock: 10 μg/mL

  Dilute to 1 ug/mL: 10 μL stock + 90 μL Assay medium.

Stock (uL)	1X Assay medium for Sytox and Caspase (µL)	Conc. in dilution plate	Final Conc. of TNF-α on cells
10 uL of 1 µg	990	10 ng/mL	5 ng/mL
5 uL of 1 µg	995	5 ng/mL	2.5 ng/mL
200 uL of 5 ng	300	2 ng/mL	1 ng/mL
100 uL of 5 ng	400	1 ng/mL	0.5 ng/mL
100 uL of 5 ng	900	500 pg/mL	250 pg/mL
200 uL of 500 pg	300	200 pg/mL	100 pg/mL
100 uL of 500 pg	400	100 pg/mL	50 pg/mL
50 uL of 500 pg	450	50 pg/mL	25 pg/mL
20 uL of 500 pg	480	20 pg/mL	10 pg/mL
10 uL of 500 pg	490	10 pg/mL	5 pg/mL
0 uL	500	0 pg/mL	0 pg/mL

For Unknown Samples: (Quantitated by Gel): TNF-a Library:

Normalize all the samples to the same starting concentration (500 ng/mL) as follows:

Neat: 500 ng /mL: 100 µL

1:10 of 500 ng = **50** ng/mL: 20 µL neat + 180 µL RPMI

1:10 of 50 ng = **5** ng/mL: 20 µL of 50 ng/mL + 180 µL RPMI

1:10 of 5 ng/mL = **0.5** ng/mL: 20 µL of 0.5 ng/mL + 180 µL RPMI

6) For Sytox assay: On a separate dilution plate, add 60 µL of each diluted sample to 60 µL of 2X Sytox assay media. Transfer 100 µL of diluted samples to the cells cultured in 100 uL media. Incubate at 37°C for 6 hrs. Read the plate using a fluorescence microplate reader with filters appropriate for fluorescein (485 nm excitation filter and 530 nm emission filter).

7) For Caspase assay: On a separate dilution plate, add 35 µL of each diluted sample to 35 µL of 2X Caspase assay media. Transfer 50 µL of dil. Samples to the cells cultured in 50 µL media. Incubate at 37°C for 4 hours. After 4 hrs. add Caspase Substrate (100 µL/well) [Predilute substrate 1:10]. Incubate 2 more hrs. at 37°C. Read (fluorescence).

C) Data Analysis:

The fluorescence signal is directly proportional to the number of apoptotic cells. Plot fluorescence vs. TNF-α standard concentration to make a standard curve. Compare the fluorescence obtained from the highest point on the standard curve (5 ng/mL) to the fluorescence obtained from the unknown samples, to determine the % activity of the samples.

The data may be analyzed using a four-parameter fit program to determine the 50% effective concentration for TNF (EC_{50}). % Activity of unknown samples = (Fluor. Of unknown samples/ fluor. of 5 ng/mL std. Point) x 100.

Example 2

PDA Calculations for soluble TNF-R (p55)

Using publicly available protein three-dimensional structures for the p55 TNFR (Protein Data Bank codes 1ext, 1ncf, 1nr) both alone and complexed with its ligand, PDAcan be used to design optimized soluble p55 receptors as TNF-α antagonists. For the library shown below, the sequences shown were generated using PDA relative to the Protein Data Bank 1 ext numbering scheme. Amino acid residues known from the structure of the receptor-TNF-α complex to be critical for p55 binding to TNFα were designed around. The results shown in Table 1 are an example of a library in which 15 positions from the wild-type p55 receptor were used for PDA design. Four of the positions chosen were nonpolar, 7 of the positions were charged, and 4 were polar. The library shown in Table 1 was pooled from five independent designs, and a 15% cutoff was applied for each position in the library. The size of the library for a single mutation is 78 and the entire library is 1.5×10^{10} sequences. The wild-type (WT) sequence is shown in the first line of the table. The mutation pattern for soluble p55 receptors at given position is shown in the remainder of the table.

Table 1.

|	54	56	57	59	62	65	67	68	69	70	95	97	98	101	103
WT	N	H	L	H	S	K	R	K	E	M	H	W	S	L	Q
	V	H	L	A	A	K	V	R	A	A	K	F	S	L	I
				T			L	K				A		K	
	E		K	E	R	R	R	D	K	M		E	T	E	F
				D	Q	Q	K		H	D	H	D	H	R	Y
				Q			E		W						K
	N	W		R	L					E					
		R		Y					S			W	W		
		K	F	K	N										R
	F	F		F					T	L		T		Q	
												K			
												Q			
					G		Q								
					S										
				H					E						Q

CLAIMS

We claim:

1. A non-naturally occurring variant TNF-α protein comprising an amino acid sequence that has at least one amino acid substitution as compared to the wild-type TNF-α sequence, wherein said variant TNF-α protein will preferentially interact with the wild-type TNF-α to form mixed trimers incapable of activating receptor signaling.

2. A non-naturally occurring TNF-α protein according to claim 1 wherein said TNF-α protein has from 3 to 5 amino acid substitutions as compared to wild-type TNF-α sequence.

3. The non-naturally occurring TNF-α protein according to claim 1, wherein said substitutions are selected from amino acid residues at positions 21, 30, 31, 32, 33, 35, 65, 66, 67, 111, 112, 115, 140, 143, 144, 146 and 147.

4. The non-naturally occurring TNF-α protein according to claim 1, wherein said substitutions are selected from the group of substitutions consisting of D143E, D143N, D143S, A145R, A145K, A145E, E146K, E146R and A84V.

5. A recombinant nucleic acid encoding the non-naturally occurring TNF-α protein of claim 1.

6. An expression vector comprising the recombinant nucleic acid of claim 5.

7. A host cell comprising the recombinant nucleic acid of claim 5.

8. A host cell comprising the expression vector of claim 6.

9. A method of producing a non-naturally occurring TNF-α protein comprising culturing the host cell of claim 7 under conditions suitable for expression of said nucleic acid.

10. The method according to claim 9 further comprising recovering said TNF-α protein.

11. A pharmaceutical composition comprising a non-naturally occurring TNF-α protein according to claim 1 and a pharmaceutical carrier.

12. A method for treating a TNF-α related disorder comprising administering a non-naturally occurring TNF-α protein to a patient.

13. The method according to claim 12, wherein said condition is rheumatoid arthritis.
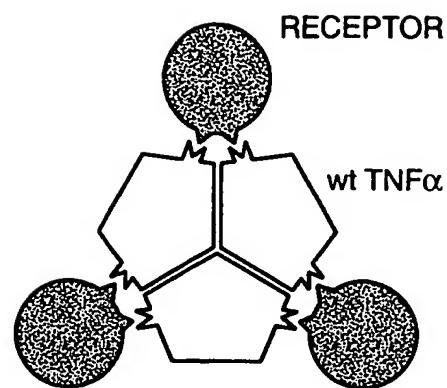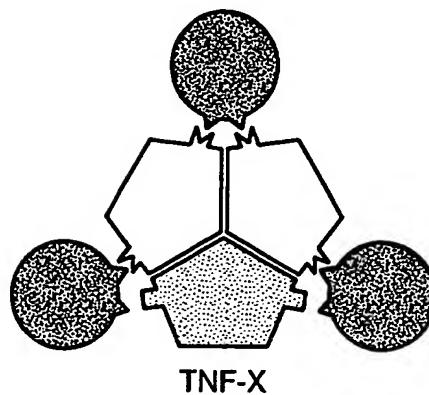
RECEPTOR

wt TNFα

FIG._1A


TNF-X

FIG._1B

TNF-TNFR TRIMER COMPLEX

SIDE VIEW

TOP VIEW

FIG._2

FIG._3

4 / 14

TNFα BINDING SITES

SITE 1

SITE 2
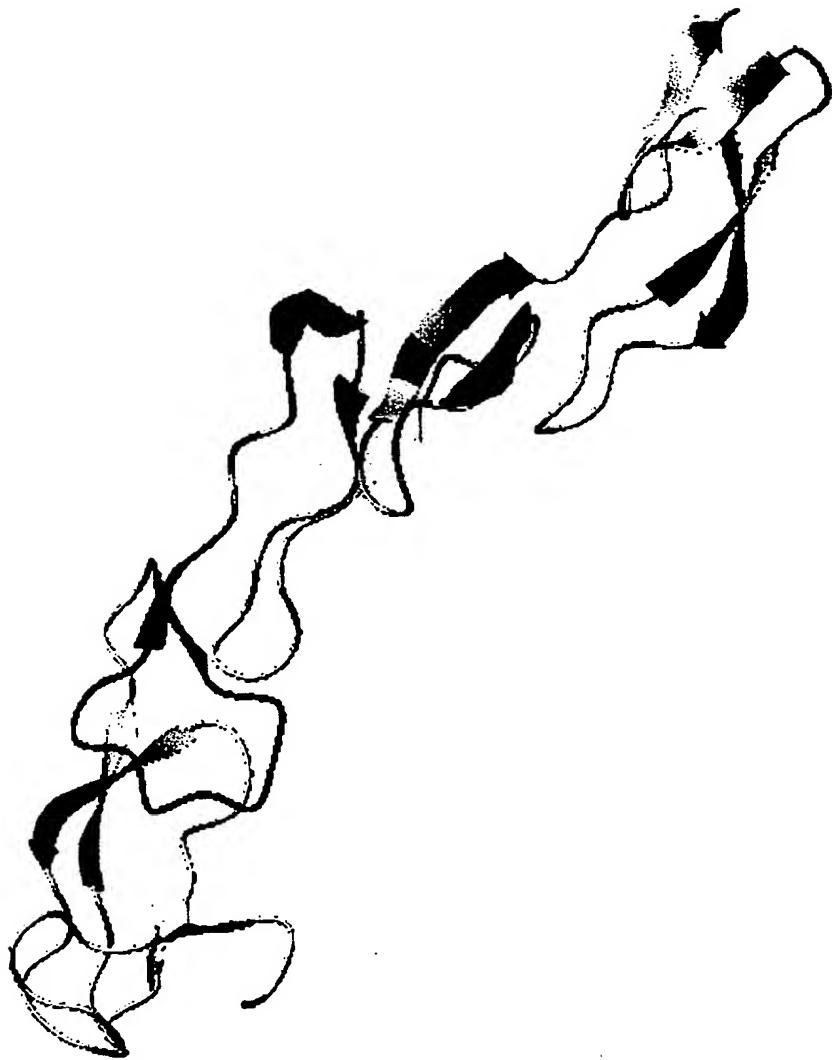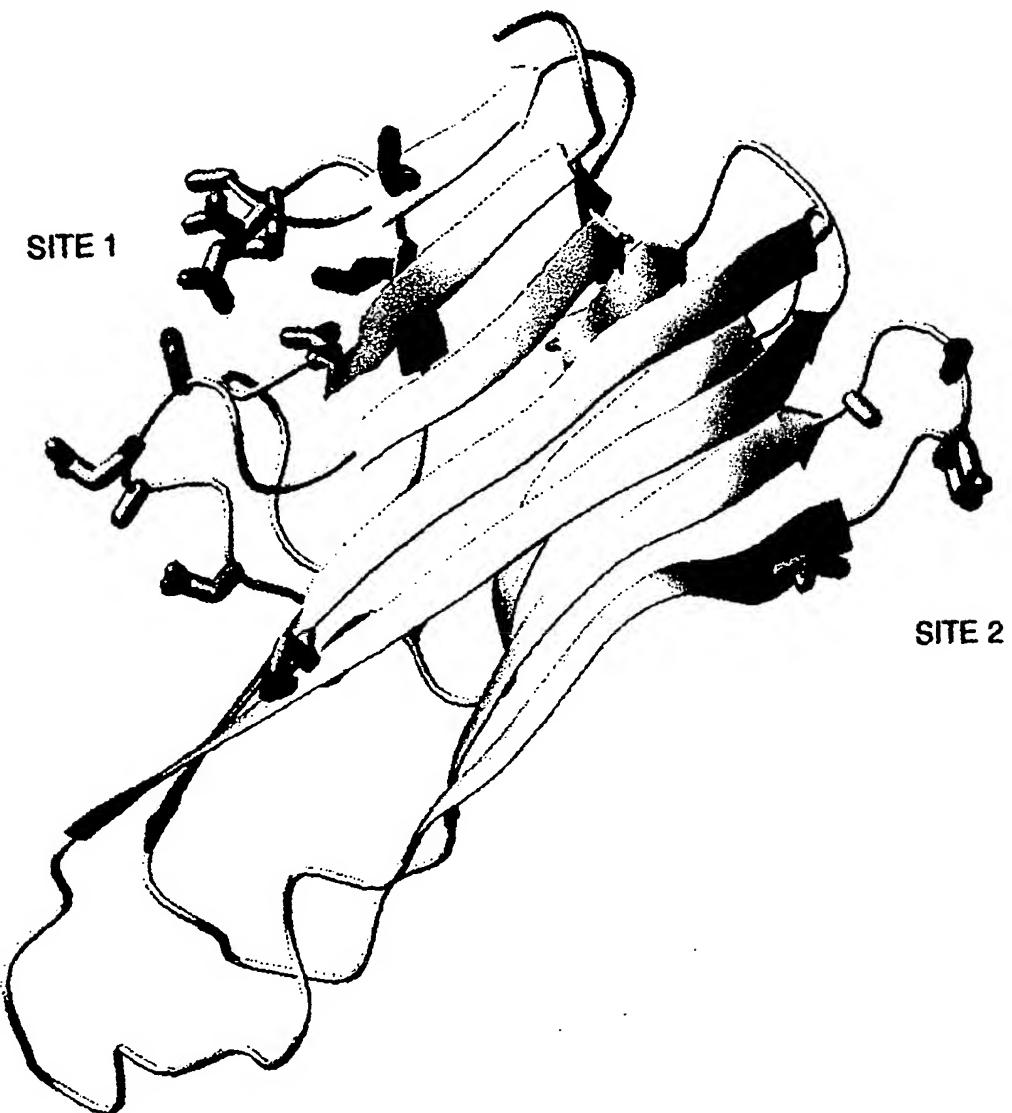
FIG._4

TNFa TRIMER INTERFACE
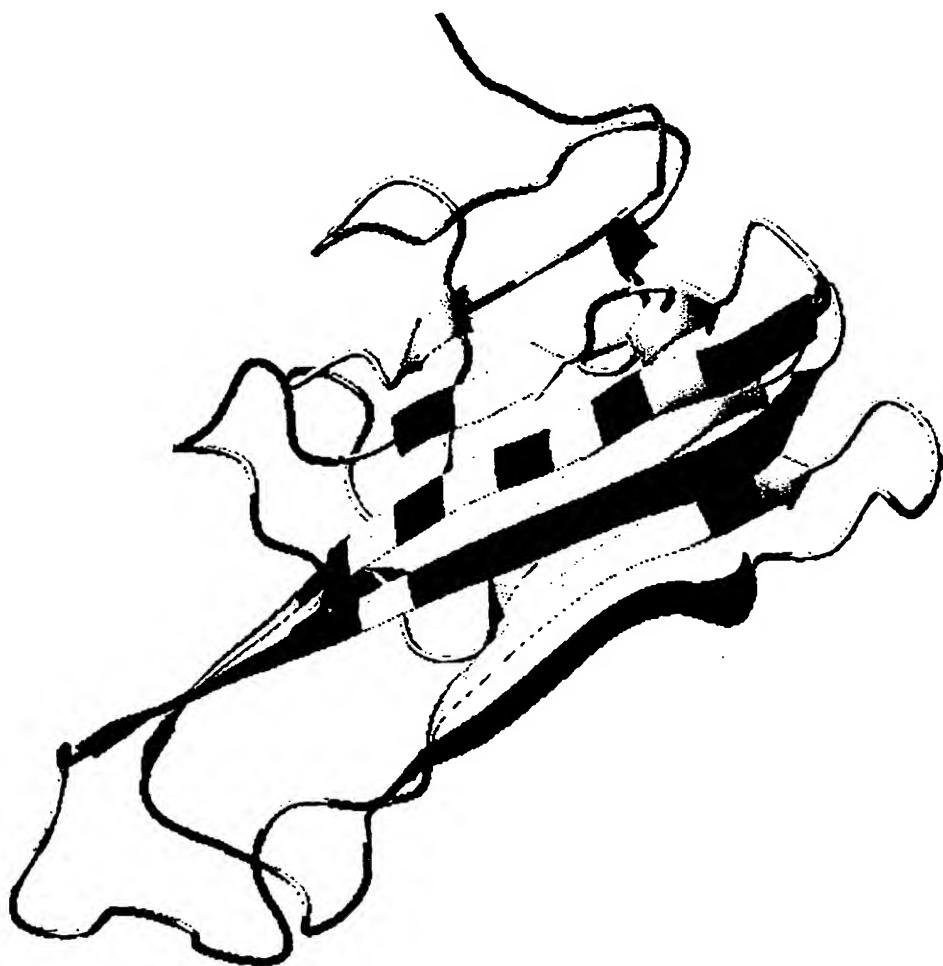
FIG._5

```
1    atgcaccacc accaccacca cgtacgctcc tcctcccgca ctccgtccga caaaccggta
61   gctcacgtag tagctaaccc gcaggctgaa ggtcagctgc agtggctgaa ccgccgcgct
121  aacgctctgc tggctaacgg tgtagaactg cgcgacaacc agctggtagt accgtccgaa
181  ggtctgtacc tgatctactc ccaggtactg ttcaaaggtc agggttgtcc gtccactcac
241  gtactgctga ctcacactat ctcccgcatc gctgtatcct accagactaa agtaaacctg
301  ctgtccgcta tcaaatcccc gtgtcagcgc gaaactccgg aaggtgctga agctaaaccg
361  tggtacgaac cgatctacct gggtggtgta ttccagctgg aaaaaggtga ccgcctgtcc
421  gctgaaatca accgcccgga ctacctggac ttcgctgaat ccggtcaggt atacttcggt
481  atcatcgctc tgtga
```

FIG._6A

```
1    MHHHHHHVRS SSRTPSDKPV AHVVANPQAE GQLQWLNRRA NALLANGVEL RDNQLVVPSE

61   GLYLIYSQVL FKGQGCPSTH VLLTHTISRI AVSYQTKVNL LSAIKSPCQR ETPEGAEAKP

121  WYEPIYLGGV FQLEKGDRLS AEINRPDYLD FAESGQVYFG IIAL
```

FIG._6B

Wild-type TNF amino acid	Wild-type TNF amino acid number	Mutants created
Q	21	R
N	30	D
R	31	I, D, E
R	32	D, E, S
A	33	E
A	35	S
K	65	D, T, M, W, I, Q, S, N, V, E
G	66	Q, K
Q	67	D, W, Y, R, K, S
A	111	R, E
K	112	D, E
Y	115	Q, K, E, N, R, F, H, M, L, I, W, D, T, S
D	140	R, K
D	143	E, N, Q, S, R, K
F	144	N
A	145	R, D, K, N, H, T, Q, E, Y, M, S, F
E	146	N, K, R, S
S	147	R

ALSO MADE DOUBLE MUTANTS K65E/D143K, K65E/D143R, K65D/D143K AND K65D/D143R

FIG._7

| Assay Done: 1.18.01-1.25.01 | | | | | | | |
| | | | | % TNF-a Activity** | | | |
Clone #	Oligo Name	Well location	Protein Conc. (ng/ul)	1:10	1:100	1:1000	1:10000
1	Q21Rf	A01	34.43	97	100	94	62
2	N30Df	A02	51.40	108	100	99	70
3	R31If	A03	46.26	91	74	64	45
4	R31Df	A04	56.82	98	92	80	58
5	R31Ef	A05	56.42	102	94	87	68
6	R32Df	A06	60.17	90	65	56	42
7	R32Ef	A07	52.39	72	44	26	21
8	R32Sf	A08	44.12	86	65	39	26
9	A33Ef	A09	46.12	64	43	25	18
10	A35Sf	A10	37.86	109	105	81	75
11	K65Df	A11	0.00	14	13	13	15
12	K65Tf	A12	49.58	104	97	80	81
13	K65Mf	B01	29.42	110	89	85	64
14	K65Wf	B02	12.73	65	36	22	17
15	K65If	B03	11.61	89	47	31	21
16	K65Qf	B04	10.45	85	46	32	19
17	K65Sf	B05	46.02	123	76	107	81
18	K65Nf	B06	8.77	87	61	77	39
19	K65Vf	B07	26.39	117	119	110	78
20	K65E	B08	0	30	14	15	13
21	G66Q	B09	16.55	122	35	19	15
22	G66Kf	B10	6.83	68	30	19	16
23	Q67Df	B11	47.94	109	97	83	64
24	Q67W	B12	0	15	14	14	14
25	Q67Yf	C01	48.57	127	112	106	85
26	Q67Rf	C02	48.40	131	121	119	95
27	Q67Kf	C03	38.83	122	126	117	87
28	Q67Sf	C04	35.72	122	118	108	86
29	A111Rf	C05	35.67	119	113	109	89
30	A111Ef	C06	37.54	122	119	104	78
31	K112Df	C07	14.06	17	15	15	14
32	K112Ef	C08	70.65	128	114	90	57
33	Y115Qf	C09	6.34	15	15	15	13
34	Y115Kf	C10	4.83	16	15	16	15
35	Y115Ef	C11	5.35	17	15	15	15
36	Y115Nf	C12	3.55	14	13	14	13
37	Y115Rf	D01	0.00	17	15	14	14
38	Y115Ff	D02	58.81	118	106	87	56
39	Y115Hf	D03	37.23	100	86	50	26

FIG._8A

	Assay Done: 1.18.01-1.25.01						
				% TNF-a Activity**			
Clone #	Oligo Name	Well location	Protein Conc. (ng/ul)	1:10	1:100	1:1000	1:10
40	Y115Mf	D04	9.56	15	14	15	14
41	Y115Lf	D05	2.16	16	14	14	15
42	Y115If	D06	17.12	20	15	14	14
43	Y115Wf	D07	75.12	113	94	70	39
44	Y115Df	D08	0.00	13	13	14	13
45	Y115Tf	D09	0.00	14	14	14	14
46	Y115Sf	D10	0.00	16	14	13	14
47	D140Rf	D11	13.32	109	100	77	44
48	D140Kf	D12	26.70	107	97	81	56
49	D143Ef	E01	35.86	25	15	13	14
50	D143Nf	E02	38.52	14	13	14	13
51	D143Qf	E03	19.28	14	13	14	13
52	D143Sf	E04	45.91	19	14	12	13
53	D143Rf	E05	0.00	13	14	14	14
54	D143Kf	E06	0.00	14	13	14	14
55	F144Nf	E07	48.24	96	98	90	83
56	A145Rf	E08	41.18	14	11	11	12
57	A145Df	E09	46.49	87	77	58	43
58	A145Kf	E10	55.47	13	12	12	12
59	A145Nf	E11	49.33	92	77	61	42
60	A145Hf	E12	0.00	13	12	12	12
61	A145Tf	F01	46.08	101	89	82	68
62	A145Qf	F02	52.05	92	78	70	56
63	A145Ef	F03	53.29	20	13	12	12
64	A145Yf	F04	35.01	39	22	15	13
65	A145Mf	F05	47.97	54	28	18	13
66	A145Sf	F06	56.30	95	86	83	76
67	A145Ff	F07	37.02	32	17	13	12
68	E146Nf	F08	32.75	84	74	53	28
69	E146Kf	F09	33.28	12	12	11	12
70	E146Rf	F10	32.22	12	11	12	12
71	E146Sf	F11	41.18	84	72	54	31
72	S147Rf	F12	5.70	54	34	23	15
73	A84V	G07	27.84	15	15	14	13
74	K65E/D143K	G1	8.39	34	18	15	13
75	K65E/D143R	G2	9.97	16	15	14	13
76	K65D/D143K	G3	0.00	15	13	14	13
77	K65D/D143R	G4	0.00	15	14	14	15
78	WT	G5	53.92	130	117	112	90

FIG._8B

% Activity is based on the. highest Std. Pt

TNF-a ACTIVITY ASSAY:
ALL THE CLONES (1.18.01 - 2.01.01)

| ♦ 1/18/01 (1:10) |
| ■ 1/25/01 |
| △ 2/1/01 |

% BASED ON HIGHEST STD. PT.

CLONE #

FIG._9A

COMPARISON OF 1/18, 1/25 AND 2/01/01 DATA
SELECTED CLONES REPEATED
FOR CONFIRMATION

| ♦ 1/18/01 |
| ■ 1/25/01 |
| △ 2/1/01 |

% BASED ON HIGHEST STD. PT.

CLONE #
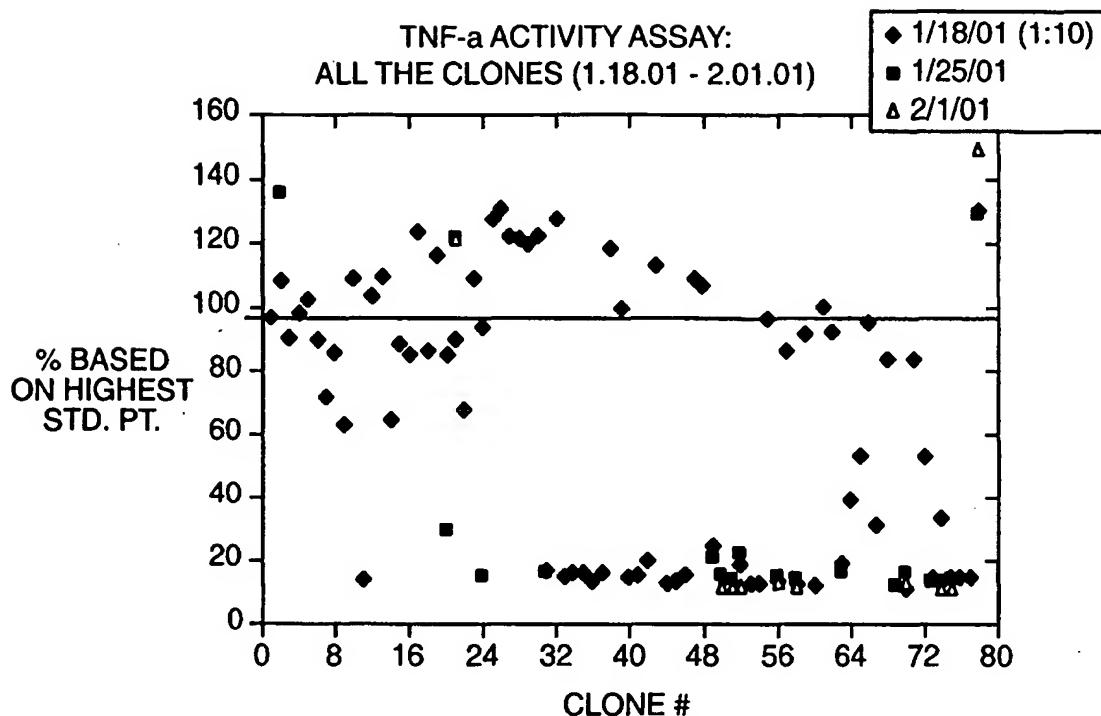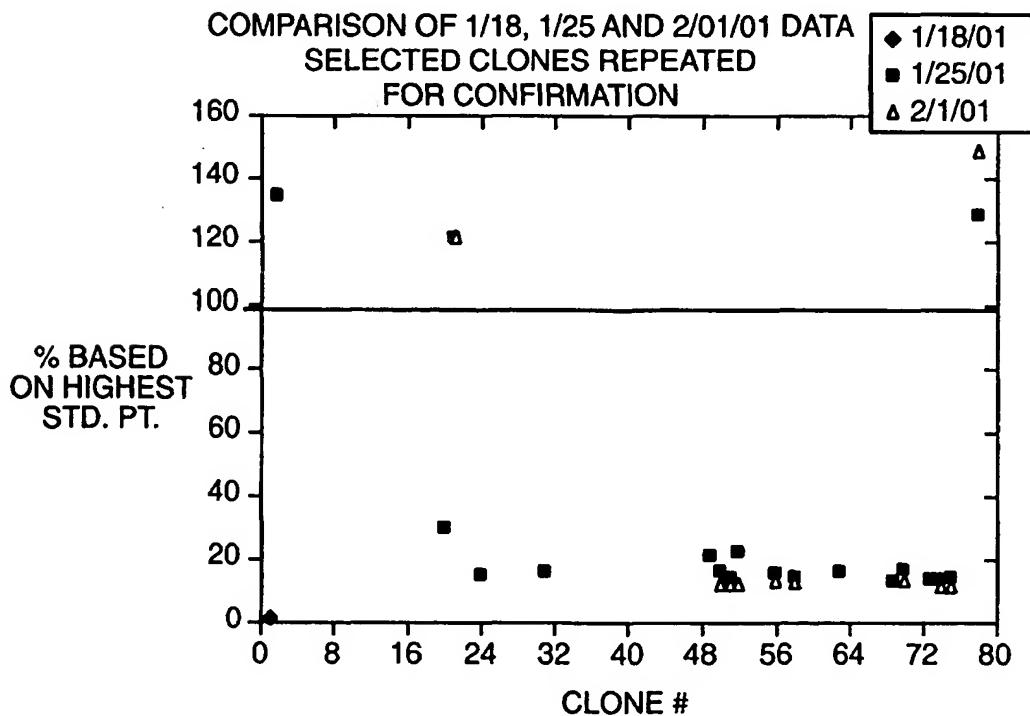
FIG._9B

WT	PDA Relative Probability Distribution
72T	Q:643 A:238 G:119
73H	E:605 Y:259 R:71 Q:58 N:2 K:2 D:2 W:1
75L	Q:424 I:269 N:170 M:116 A:15 G:6
86S	K:475 R:333 T:77 S:55 D:32 Q:15 E:13
87Y	D:402 N:170 R:108 Y:62 V:31 E:31 H:31 F:30 M:26 K:26 L:23 Q:18 T:18 S:12 A:9 G:3
97I	K:659 Q:187 R:104 N:21 M:9 E:7 V:5 T:4 I:2 D:2
137N	D:687 Q:75 E:62 M:48 K:43 L:37 R:33 S:9 N:3 A:3

FIG._10

WT	PDA Relative Probability Distribution
Q21	R1000
N30	D1000
R31	I1000
R32	H1000
A33	E1000
A35	S1000
K65	R585 D146 K110 T42 M27 W15 I15 Q10 S9 N9 V1
G66	Q813 K187
Q67	D623 W209 Y83 R43 H31
A111	R959 E41 K41 S1
K112	K1000
Y115	Q230 K154 E116 N84 Y81 R72 F69 H43 M39 L36 I26 W25 D11 T8 S6
D140	D1000
L143	D680 E130 N110 Q33 S29 R12 K6
F144	F695 N305
A145	R456 D196 K124 N76 H67 T43 Q25 E9 Y1 M1 S1 F1
E146	N489 K377 R111 D12 S10 E1
S147	R1000

FIG._11

TRAF2(310-) DQDKIEALSSKVQQLERSIGLKDLAMADLEQKVLEMEA STYDG

FIG._12A

TRAF3(374-) VARNTGLLESQLSRHDQMLSVHDIRLADMDLRFQVLET ASYNG

FIG._12B

TRAF5(343-) NDQRLAVLEEETNKHDTHINIHKAQLSKNEERFKLLEG TCYNG

FIG._12C

TRAF1(225-) DRERILSLEQRVVELQQTLAQKDQALGKLEQSLRLMEE ASFDG

FIG._12D

TRAF6(309-) QDHQIRELTAKMETQSMYVSELKRTIRTLEDKVAEIEA QQCNG

FIG._12E

TRAF4(201-) --------------CALVSRQRQELQELRRELEELSV GS-DG

FIG._12F

FIG._13

FIG._ 14

BLACK BOX =
REGION TO
BE MUTATED
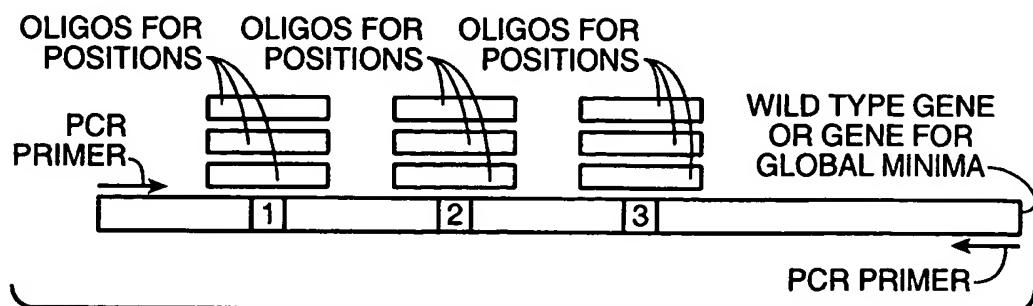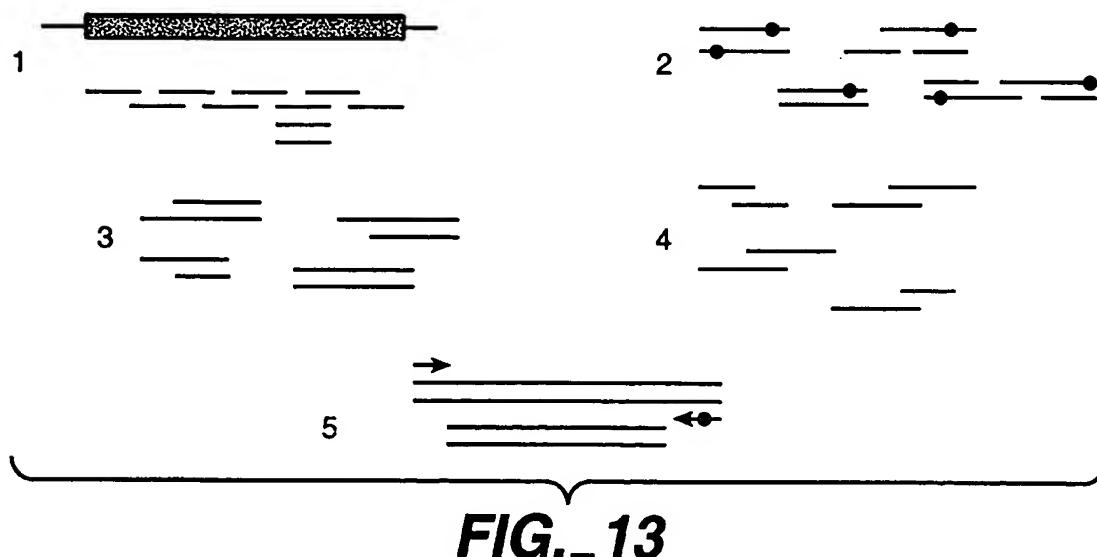
F1   F2   F3   TEMPLATE
                                    DNA
R1   R2   R3
                R4
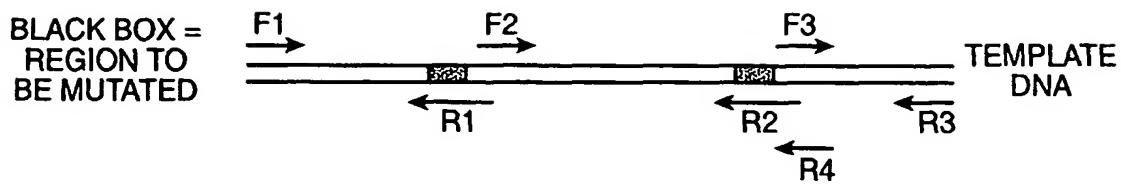
STEP 1:   SET UP 3 PCR REACTIONS:

          PRODUCTS:

TUBE 1:

TUBE 2:

TUBE 3:

STEP 2:   SET UP PCR REACTION WITH PRODUCTS OF TUBE 1 +
          PRODUCTS TUBE 2 + F1 + R4.

          HEAT + ANNEAL PHASE OF PCR,

F1

                                    R4

          SYNTHESIS PHASE OF PCR,

F1   AMPLIFICATION PHASE DURING
                                    SUBSEQUENT CYCLES USING F1 + R4.

                                    R4
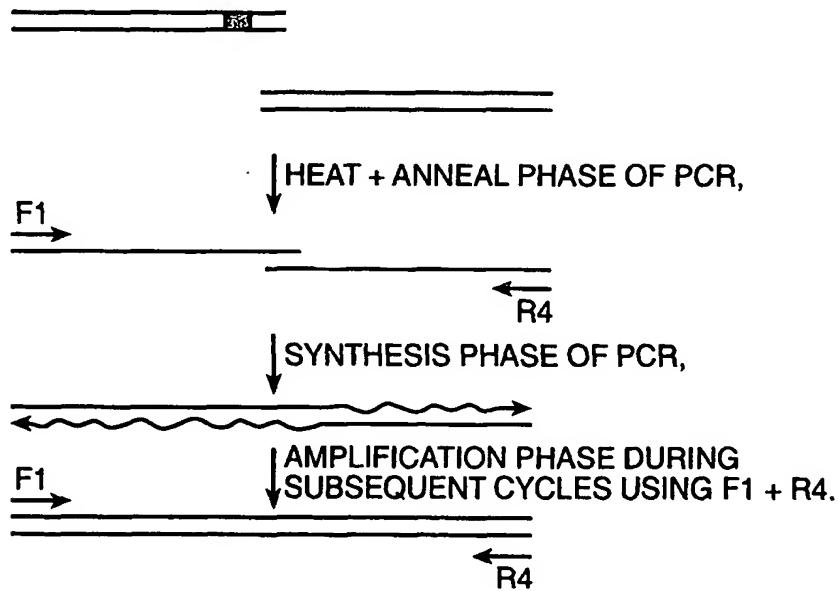
STEP 3:   REPEAT STEP 2 USING PRODUCT FROM STEP 2 + PRODUCT
          FROM STEP 1, TUBE 3 + PRIMERS F1 + R3.

FIG._15

SUBSTITUTE SHEET (RULE 26)
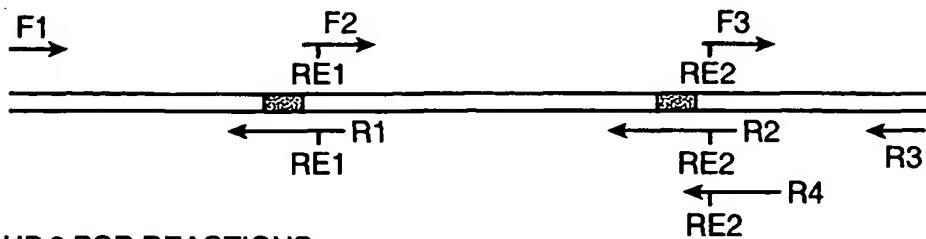
STEP 1: SET UP 3 PCR REACTIONS:

TUBE 1:

TUBE 2:

TUBE 3:

STEP 2: DIGEST PRODUCTS FROM STEP 1 WITH SUITABLE
RESTRICTION ENDONUCLEASES.

STEP 3: LIGATE DIGESTED PRODUCT FROM STEP 2, TUBE 2 WITH
DIGESTED PRODUCT FROM STEP 2, TUBE 1.
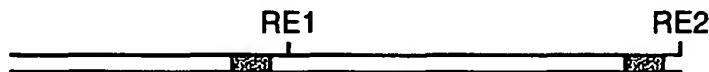
STEP 4: AMPLIFY VIA PCR LIGATED PRODUCTS OF STEP 3 WITH F1 + R4.
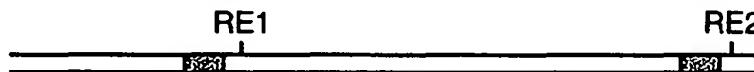
STEP 5: DIGEST AMPLIFIED PRODUCT OF STEP 4
WITH RESTRICTION ENDONUCLEASE #2.

STEP 6: LIGATE PRODUCT FROM STEP 5 WITH
PRODUCT FROM STEP 2, TUBE 1.
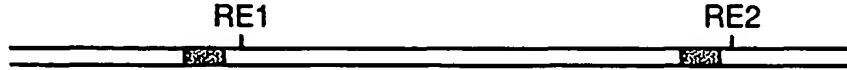
STEP 7: AMPLIFY PRODUCT FROM STEP 6 WITH F1 + R3.

FIG._16

DIAGRAM 3

FIG._17